

10E17

SEARCH REQUEST FORM

Requestor's Name: D. B. Jahan Serial Number: 08/483,534
Date: 04/16/98 Phone: 305-2150 Art Unit: 1646

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

Please search

AA 1-168 SEQ ID NO: 2 NA 94-597 SEQ ID NO: 1

" 1-307 SEQ ID NO: 2 " 1-597 SEQ ID NO: 1

only goes to 168
Thanks.

STAFF USE ONLY

Date completed: 4-28-98
Searcher: MANC
Terminal time: 15
Elapsed time: 15
CPU time: 5
Total time: 20
Number of Searches: 1
Number of Databases: 7

Search Site

☒ STIC
☒ CM-1
☐ Pre-S

Type of Search

☒ N.A. Sequence
☒ A.A. Sequence
☐ Structure
☐ Bibliographic

Vendors

☒ STIC
☒ STN
☒ Dialog
☐ APS
☐ Geninfo
☐ SDC
☐ DARC/Questel
☐ Other

THIS PAGE BLANK (USPTO)


```

QY 3 VIPSRLDIRVGIITVEKHPDADSLYEKIDVGEAEPRIVSGLVGFVPEKELQDRLLVV 62
DB 208 LCNLKPAPKMGVLSQAMVWCAS--SPEK-VEILAPNGSVGDRITFDAPP-GEPPKELN 263
QY 63 LCNLKPAPKMGVLSQAMVWCAS--SPEK-VEILAPNGSVGDRITFDAPP-GEPPKELN 122
DB 264 PKKIMEOIOPLDHTNAECVATYKAPFEVK-GKVCRA 301
QY 123 PKKVEFEKIQADFEKISECIAOMKOTNFMTKLGISCSKS 161

RESULT 2
ID Q20970 PRELIMINARY; PRT; 917 AA.
AC Q20970;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
OS F5883.5.
OC CAENORHABDITIS ELEGANS.
RN EUKARYOTA; METAZOA; ACLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RP SEQUENCE FROM N.A.
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RA BONTLED J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,
RA THIRRY-MING J., THOMAS K., VAIDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;
RA NATURE 368:32-38(1994).
DR EMBL; Z73427; E244446;
SQ SEQUENCE 917 AA; 101713 MW; D148DA58 CRC32;

Query Match 49.7%; Score 595; DB 3; Length 917;
Best Local Similarity 51.5%; Pred. No. 7,72e-103;
Matches 87; Conservative 39; Mismatches 37; Indels 6; Gaps 5;

DB 752 DDTIDVGRDLMRGRIKCEHPDADSLYEKIDVGEAEPRIVSGLVGFVPEKELQDRLL 111
QY 1 EEVYIP-SRDIRGKIITVEKHPDADSLYEKIDVGEAEPRIVSGLVGFVPEKELQDRLL 59
DB 812 VVVLNKLKPAKMGVLSQAMVWCAS--SPDK-VEIMEVPADSKPGPPVVCPPYTH-RPDE 867
QY 60 VVVLNKLKPAKMGVLSQAMVWCAS--SPDK-VEIMEVPADSKPGPPVVCPPYTH-RPDE 111
DB 868 QLNPKKKIWTVAEDKLVSAGFAEKMGQPLLGSESKMPTPLRGVH 916
QY 120 ELKPKKKVEKIQADFEKISECIAOMK-OTNFMTKLGISCSKSLKGN 167

RESULT 3
ID Q12904 PRELIMINARY; PRT; 312 AA.
AC Q12904;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
RN EUTHERIA; PRIMATES.
RP SEQUENCE FROM N.A.
RL MEDLINE; 95014290.
RA KAO J., HOCK K., PAN Y., HAHNEL I., LIBUTTI S.K., KAYTON M.L.,
RA GRINSCHETT T., CHABOT J., NOWYGRON R., GREENBERG S., KUANG W.J.,

```

```

RA LEUNG D.W., HAYWARD J.R., KISIEL W., HEATH M., BRETT J.,
RA STERN D.M.;
RL J. BIOL. CHEM. 269:25106-25119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94193665.
RA KAO J., HOCK K., KISIEL W., SELJELID R., BURNIER J., STERN D.;
RA J. BIOL. CHEM. 269:9774-9782(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93015897.
RA KAO J., RYAN J., BRETT G., CHEN J., SHEN H., PAN Y., GODMAN G.,
RA FAMILLETI P.C., WANG F., PAN Y.E., STERN D., CLAUS M.;
RL J. BIOL. CHEM. 267:20239-20247(1992).
DR EMBL; U10117; G498910;
FT CHAIN 147 312 ENDOTHELIAL-MONOCYTE ACTIVATING
POLYPEPTIDE II.
SQ SEQUENCE 312 AA; 34326 MW; D012B77F CRC32;

Query Match 49.1%; Score 588; DB 2; Length 312;
Best Local Similarity 53.8%; Pred. No. 2,76e-101;
Matches 84; Conservative 32; Mismatches 35; Indels 5; Gaps 4;

DB 153 SRLDIRGKIITVEKHPDADSLYEKIDVGEAEPRIVSGLVGFVPEKELQDRLLVV 212
QY 6 SRLDIRGKIITVEKHPDADSLYEKIDVGEAEPRIVSGLVGFVPEKELQDRLLVV 65
DB 213 LKAPKRGVLSQAMVWCAS--SPEK-TEILAPNGSVGDRITFDAPP-GEPPKELNPK 268
QY 66 LKAPKRGVLSQAMVWCAS--SPEK-TEILAPNGSVGDRITFDAPP-GEPPKELNPK 125
DB 269 KIMEOIOPLDHTNAECVATYKAPFEVK-GKVCRA 303
QY 126 KVEKIQADFEKISECIAOMKOTNFMTKLGISCSKS 161

RESULT 4
ID Q28819 PRELIMINARY; PRT; 658 AA.
AC Q28819;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
OS METHONYL-TRNA SYNTHETASE (METS).
GN AF1453.
OC ARCHAEOLIBUS FUGIDIOS.
RN ARCHAEOLIBUS FUGIDIOS.
RP SEQUENCE FROM N.A.
RA KLEIN H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E., KETCHUM K.A.,
RA DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L.,
RA KERLAVAGE A.R., GRAHAM D.E., KIRPDES N.C., FLEISCHMANN R.D.,
RA DOUGHERTY B.A., LEE N.H., SUTTON G.G., GILL S., KIRKNESS E.F.,
RA DOUGHERTY J., MCKENNEY K., ADAMS M.D., LOFTUS B., PETERSON S.,
RA REICH C.I., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L., OVERBECK R.,
RA GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTERBACK T., COTTON M.D.,
RA SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M., SADOW P.W.,
RA D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M.,
RA OLSEN G.J., FRASER C.M., SMITH H.O., WOSE C.R., VENTER J.C.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA KLEIN H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E., KETCHUM K.A.,
RA DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L.,
RA KERLAVAGE A.R., GRAHAM D.E., KIRPDES N.C., FLEISCHMANN R.D.,
RA DOUGHERTY B.A., LEE N.H., SUTTON G.G., GILL S., KIRKNESS E.F.,
RA DOUGHERTY J., MCKENNEY K., ADAMS M.D., LOFTUS B., PETERSON S.,
RA REICH C.I., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L., OVERBECK R.,
RA GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTERBACK T., COTTON M.D.,
RA SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M., SADOW P.W.,
RA D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M.,
RA OLSEN G.J., FRASER C.M., SMITH H.O., WOSE C.R., VENTER J.C.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

```


DR EMBL; AE001003; G2649118; -
 KW AMINOACYL-TRNA SYNTHASE.
 SQ SEQUENCE 658 AA; 75695 MW; 7761EB99 CRC32;

Query Match 22.0%; Score 263; DB 9; Length 658;
 Best Local Similarity 44.1%; Pred. No. 1,34e-31;

Matches 45; Conservative 19; Mismatches 32; Indels 6; Gaps 5;

DB 561 KIDIRIGRVLKAEKVKRSKLKLIKIDIDG-EDQROVSGIAEDYTP-ELEGKLVVLAN 618
 QY 7 RDIIRGKIITVEKHPDADSLYERKIDVGEAEPRVYSGIVQ-FVPEKELODRLVVLCN 65

DB 619 LKPAFMGVESGGMILAEKDG-KAV-LITPEKEVEPSTRV 657
 QY 66 LKPKMRGVESOGMLLCASIEGINROVEPLDPAGSAPGEHV 107

RESULT 5 PRELIMINARY; PRT; 629 AA.

ID 033925;
 AC 033925;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE METHIONYL-TRNA SYNTHETASE.
 OS THERMOTOGA MARITIMA.
 OC PROKARYOTA; NOT YET CLASSIFIED.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MSB8;
 RX MEDLINE; 97454499.
 RA GUTPAUD O., MARGUET E., NOLL K.M., DE LA TOUR C.B., FORTERRE P.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 94:10606-10611(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MSB8;
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U76417; G2645225; -
 KW AMINOACYL-TRNA SYNTHASE.
 SQ SEQUENCE 629 AA; 73004 MW; 7209A0F4 CRC32;

Query Match 19.7%; Score 236; DB 9; Length 629;
 Best Local Similarity 45.5%; Pred. No. 3,34e-26;
 Matches 35; Conservative 21; Mismatches 18; Indels 3; Gaps 3;

DB 531 SKVDRIANVLAEKVPNSRKLRLIIDIG-TEKROIVAGIAEHYR-ELVYKLVVVA 588
 QY 6 SRLDIRVGKIITVEKHPDADSLYERKIDVGEAEPRVYSGIV-QFVPEKELODRLVVLC 64

DB 589 NLKPAKMGIESOGML 605
 QY 65 NLKPKMRGVESOGML 81

RESULT 6 PRELIMINARY; PRT; 651 AA.

ID 026687;
 AC 026687;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE METHIONYL-TRNA SYNTHETASE.
 GN MTH587.
 OS METHANOBACTERIUM THERMOAUTOTROPHICUM.
 OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOBACTERIALES;
 OC METHANOBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
 RA ALDEDEDE T., BASHIRZADER R., BLAKELY D., COOK R., GILBERT K.,
 RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTIER B., QIU D.,
 RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
 RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,

RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
 RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
 RL J. BACTERIOL. 179:7135-7155(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RA SMITH D.R.;
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AE000841; G2621666; -
 KW AMINOACYL-TRNA SYNTHASE.
 SQ SEQUENCE 651 AA; 74631 MW; C8E64F97 CRC32;

Query Match 18.5%; Score 222; DB 9; Length 651;
 Best Local Similarity 44.7%; Pred. No. 1,88e-23;
 Matches 42; Conservative 20; Mismatches 27; Indels 5; Gaps 5;

DB 558 MDIRGVIRSAERIGSDKLKLIKIDGGERMO-VVAGIAEKYSP-EDLVERKITVNL 615
 QY 8 LDIRVGKIITVEKHPDADSLYERKIDVGEAEPRVYSGIVQ-FVPEKELODRLVVLCN 66

DB 616 KPAKLGVSSEGMVL-ANGESINI-LDPGDAEYV 647
 QY 67 KPQKMRGVESOGMLLCASIEGINROVEPLDPAG 100

RESULT 7 PRELIMINARY; PRT; 246 AA.

ID 027056;
 AC 027056;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE CONSERVED PROTEIN.
 GN MTH975.
 OS METHANOBACTERIUM THERMOAUTOTROPHICUM.
 OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOBACTERIALES;
 OC METHANOBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
 RA ALDEDEDE T., BASHIRZADER R., BLAKELY D., COOK R., GILBERT K.,
 RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTIER B., QIU D.,
 RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R., JIWANI N.,
 RA CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S., MCDUGALL S.,
 RA SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M., DANIELS C.J.,
 RA MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
 RL J. BACTERIOL. 179:7135-7155(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RA SMITH D.R.;
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AE000871; G2622074; -
 SQ SEQUENCE 246 AA; 27540 MW; A6538793 CRC32;

Query Match 11.5%; Score 138; DB 9; Length 246;
 Best Local Similarity 36.0%; Pred. No. 5,90e-08;
 Matches 31; Conservative 23; Mismatches 23; Indels 9; Gaps 6;

DB 142 VDIIRGEVNSVGNHVPADLHVCNAGGRSIRKVTIND-D-V-REN-DHYAVAL--LP 194
 QY 8 LDIRVGKIITVEKHPDADSLYERKIDVGEAEPRVYSGIVQ-FVPEKELODRLVVLCN 67

DB 195 PONEGVTSSEGMFL--GVEGVLRVDV 218
 QY 68 POKMRGVESOGMLLCASIEGINROYE 93

RESULT 8 PRELIMINARY; PRT; 201 AA.
 ID 034943;
 AC 034943;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE YTPR.
 GN BACILLUS SUBTILIS.
 OS BACILLUS SUBTILIS.
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 AN [1]
 RP SEQUENCE FROM N.A.
 RA LAPIDUS A., GALLERON N., SOROKIN A., EHRLICH D.;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 AAEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
 RA BORRIS R., BOUSIER L., BRANS A., BRAUN M., BRIGELL S.C., BRON S.,
 RA BROUILLET S., BUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMINGS N.J., DANIEL R.A.,
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,
 RA ENRIAN K.D., ERINGTON J., FABRET C., FERRARI E., FOULGER D.,
 RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALLIZI A., GALLERON N.,
 RA GIM S.Y., GLASER P., GOFEAU A., GOLIGHTLY E.J., GRANDI G.,
 RA GIUSEPPI G., GUY B.J., HAGA K., HAICH J., HARWOOD C.R., HENAUT A.,
 RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
 RA JORIS B., KARAMATA D., KASAHARA Y., KLERR-BLANCHARD M., KLEIN C.,
 RA KOBAYASHI I., KOTTER P., KONINGSTEIN G., KROCH S., KUMANO M.,
 RA KORITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 RA MEDINA N., MELADO R.P., MIZUNO M., MOSTEL D., NAKAI S., NOBACK M.,
 RA MOONE D., O'REILLY M., OGAMA K., OGIMARA A., OUDGA B., PARK S.H.,
 RA PARO V., POHL T.M., PORTETELLE D., POROLIK S., PRESCOTT A.M.,
 RA PRESEAN E., PUTIC C., PORNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
 RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADIE Y.,
 RA SAKTO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFONE F.,
 RA SERIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDO B.,
 RA SOROKIN A., TACCONTI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
 RA TAKERCHI M., TANAKOSHI A., TANAKA T., TERESTRA P., TOGNONI A.,
 RA TOSATO V., UCHITAMA S., VANDENBOL M., VANIER F., VASSAROTTI A.,
 RA VIARI A., WAMBOYT R., WEDLER E., WEDLER H., WEITZENEGER T.,
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUNO K., YATA K.,
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
 RL NATURE 390:249-256(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: Z99119; E1185855; -;
 SO SEQUENCE 201 AA; 21696 MW; E012D585 CRC32;
 Query Match 9.0%; Score 108; DB 9; Length 201;
 Best Local Similarity 35.9%; Pred. No. 4.08e-03;
 Matches 28; Conservative 15; Mismatches 31; Indels 4; Gaps 3;
 Db 98 VGVYSEKHEKHNADKLSVCKVYNGVEETLQIVGAPNDQGGKVVYAKVAMPGLVIKD 157
 Qy 12 VGIITFEKHEKHNADKLSVCKVYNGVEETLQIVGAPNDQGGKVVYAKVAMPGLVIKD 157
 Db 158 AELRGVSSGMI-CSAKE 174
 Qy 69 QKMGVSGMGLCASIE 86
 RESULT 9 PRELIMINARY; PRT; 336 AA.
 AC 028306;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN AFI973.
 OS ARCHAEOGLOBUS FULGIDUS.
 * ARCHAEBACTERIA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE.

RN [1]
 RP SEQUENCE FROM N.A.
 RA KLEIN H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E., KETCHUM K.A.,
 RA DODSON R.J., GINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L.,
 RA KERLAVAGE A.R., GRAHAM D.E., KIRPIDS N.C., FLEISCHMANN R.D.,
 RA QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S., KIRNESS E.F.,
 RA DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B., PETERSON S.,
 RA REICH C.I., MCNEIL L.R., BADGER J.H., GLODEK A., ZHOU L., OVERBECK R.,
 RA GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTERBACK T., COTTON M.D.,
 RA SPRIGGS T., ARTACH P., KAINE B.P., SYKES S.M., SADOW P.W.,
 RA D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M.,
 RA OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA KLEIN H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E., KETCHUM K.A.,
 RA DODSON R.J., GINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L.,
 RA KERLAVAGE A.R., GRAHAM D.E., KIRPIDS N.C., FLEISCHMANN R.D.,
 RA QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S., KIRNESS E.F.,
 RA DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B., PETERSON S.,
 RA REICH C.I., MCNEIL L.R., BADGER J.H., GLODEK A., ZHOU L., OVERBECK R.,
 RA GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTERBACK T., COTTON M.D.,
 RA SPRIGGS T., ARTACH P., KAINE B.P., SYKES S.M., SADOW P.W.,
 RA D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M.,
 RA OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AE00967; G2648568; -;
 SO SEQUENCE 336 AA; 38113 MW; B5BD46C3 CRC32;
 Query Match 8.3%; Score 100; DB 9; Length 336;
 Best Local Similarity 27.6%; Pred. No. 6.26e-02;
 Matches 16; Conservative 16; Mismatches 26; Indels 0; Gaps 0;
 Db 203 IPVHKKARIVSIEINPEYKLLKLNKLVNGNIVPLISQGFYTPGVADRVYM 260
 Qy 4 IPSRLDVRGKIVKHPADSLYKIDVGEKERTVSGLVQFVPEKLELDRLV 61
 RESULT 10 PRELIMINARY; PRT; 395 AA.
 AC 052942;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE FLAA, FLAB, FLAD, AND FLAC GENES, COMPLETE COS.
 GN FLAD.
 OS RHIZOBIUM MELILOTI, AND SINORHIZOBIUM MELILOTI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
 CC RHIZOBIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RU11001;
 RA PLATZER J., SCHMITT R.;
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RU11001;
 RX MEDLINE: 97474264.
 RA PLATZER J., STERR W., HAUSMANN M., SCHMITT R.;
 RL J. BACTERIOL. 179:6391-6399(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RU11001;
 RA PLATZER J., SCHMITT R.;
 RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RU11001;
 RA SCHMITT R.;
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: L76864; G1226187; -;
 DR EMBL: L49337; G2275169; -;
 SO SEQUENCE 395 AA; 40955 MW; E9A80973 CRC32;

RP SEQUENCE FROM N.A.
RX MEDLINE: 91175935
RA BRAKHAGE E.W., WOZNAT M., PUTZER H.;
RL BIOCHIMIE 72:725-734(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA WIPAT A., CARTER N., BRIGNELL C.S., GUY J.B., PIPER K., SANDERS J.,
RA EMERSON P.T., HARMOD C.R.;
RL MICROBIOLOGY 142:3067-3078(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA GOSTHEL S.F., SCHMID R., WIPAT A., CARTER N.M., EMERSON P.T.,
RA HARMOD C.R., MARAHIEL M.A.;
RL EUR. J. BIOCHEM. 244:59-65(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STAIN-168;
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERNERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA BORRIS R., BOUSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMINGS N.J., DANIEL R.A.,
RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,
RA ENRIAN K.D., ERRINGTON J., FABBET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALLIZZI A., GALLEON N.,
RA GHM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUSEPEI G., GUY B.J., HAGA K., HAIRCH J., HARMOD C.R., HENAT L.,
RA HILBERT H., HOLAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES A.,
RA JONIS B., KANAMATA D., KASAHARA Y., KLAERR-BLANCARD M., KLEIN C.,
RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROSH S., KUMANO M.,
RA KUNITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGER C.,
RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA NOONE D., O'REILLY M., OGAMA K., OGIMARA A., OUDSGA B., PARK S.H.,
RA PARO V., POHL T.M., PORTELELLA D., PORNOILLIK S., PRESCOTT A.M.,
RA PRESECAN E., PUTIC P., PORNELLE B., RAPPOPORT G., REY M., REYNOLDS S.,
RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADIY Y.,
RA SAO T., SCALANI E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA SEKIGUCHI J., SEKOWSKA A., SEOR S.J., SERRO P., SHIN B.S., SOLDO B.,
RA SOEROKIN A., TACCINI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA TAKEICHI M., TANAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA TOSATO V., UCHITAMA S., VANDEBOL M., VANNIER P., VASSAROTTI A.,
RA VIARI A., WABUTT R., WEDLER E., WEDLER H., WEITENEGGER T.,
RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
RL NATURE 390:249-256(1997).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLANILANINE + TRNA(PHE) = AMP +
CC DIPHOSPHATE + L-PHENYLANILANYL-TRNA(PHE).
DR EMBL: 275208; E1165325; -;
DR EMBL: 299118; E1184112; -;
KW AMINOACYL-TRNA SYNTHASE; LIGASE.
SQ SEQUENCE 804 AA; 87945 MW; 25395E1A CRC32;

RESULT	15	
ID	029316	PRELIMINARY; PRT; 1036 AA

	Query Match	8.1%	Score 97	DB 9	Length 804
Best Local Similarity	33.7%		Pred. No. 1.69e-01		
Matches	29	Conservative	16	Mismatches 33	Indels 8
					Gaps 8
Db	46	VIGIHLEREDHPNADRNKCTLVADIGAEAPQITCG-APNYDKGOKYAVATVGAALDPNE	104		
	:	:	:	:	:
OY	10	IRVGIITVEHPDDSDSLYEKIDVGEAPRTVSGLVQFVPEKE-LQDRLV-VILC-NL	66		
	:	:	:	:	:
Db	105	KIKRAKLTGEESNGMT-CSLQELGTE	129		
	:	:	:	:	:
OY	67	K-PO-KMKGVSQGMILCASTE-GIN	89		
	:	:	:	:	:

AC	029316;	(TREMBL:REL_05, CREATED)
DT	01-JAN-1998	(TREMBL:REL_05, LAST SEQUENCE UPDATE)
DT	01-Jan-1998	(TREMBL:REL_05, LAST ANNOTATION UPDATE)
DE	01-Jan-1998	(TREMBL:REL_05, LAST ANNOTATION UPDATE)
DI	HYPOTHETICAL 112.4 KD PROTEIN.	
GN	AF0946.	
OS	ARCHAEOGLOBUS FULIGIDUS.	
OC	ARCHABACTERIA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	KLEINK H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E.,	
RA	KETCHUM K.A., DODSON R.J., GINN M., HICKEY E.K., PETERSON J.D.,	
RA	RICHARDSON D.L., KERLAUGE A.R., GRAHAM D.E., KYRPIDES N.C.,	
RA	FELDSCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,	
RA	KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,	
RA	PETERSON S., REICH C.I., MCNEIL L.R., BADGER J.H., GLOER A., ZHOU L.,	
RA	OVERBEER R., GOCAVNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,	
RA	COTTON M.D., SPRIGGS T., ARTACH P., KAINE B.P., SYES S.M.,	
RA	SADOM P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,	
RA	MAISON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,	
RA	VENTER J.C.;	
RL	SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RA	KLEINK H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E.,	
RA	KETCHUM K.A., DODSON R.J., GINN M., HICKEY E.K., PETERSON J.D.,	
RA	RICHARDSON D.L., KERLAUGE A.R., GRAHAM D.E., KYRPIDES N.C.,	
RA	FELDSCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,	
RA	KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,	
RA	PETERSON S., REICH C.I., MCNEIL L.R., BADGER J.H., GLOER A., ZHOU L.,	
RA	OVERBEER R., GOCAVNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,	
RA	COTTON M.D., SPRIGGS T., ARTACH P., KAINE B.P., SYES S.M.,	
RA	SADOM P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,	
RA	MAISON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,	
RA	VENTER J.C.;	
RL	SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.	
DR	HYEML: AE001039: G3649670: -	
QO	HYPOTHETICAL PROTEIN.	
QO	SEQUENCE 1036 AA: 112426 MW: F0FDABF4 CRC32,	

	Query Match	7.8%	Score 94;	DB 9;	Length 1036;
	Best Local Similarity	46.7%	Pred. No.	4,45e-01;	
Matches	14;	Conservative	9;	Mismatches 6;	Indels 1; Gaps 1
Db	357	PEGYIVAKIDVGGEGELRERATNAVQVVKPK	386		
Oy	24	ADSLVEKIDVGEAEPRIVSGLVQEF-PK	52		

Search completed: Sat Apr 18 02:04:17 1998
Job time : 104 secs.

[illegible]

RESULT	15	
ID	029316	PRELIMINARY; PRT; 1036 AA

Best Local Similarity 50.0%; Pred. No. 1.59e-75;
Matches 68; Conservative 26; Mismatches 38; Indels 4; Gaps 3;

Db 206 PSAIDVGVGQAKIKHPDADSLVSTVDGEGPRVSGSLKHPFLDMQERIVVYV 265
QY 5 PSRLDIRGKIITVEKHPDADSLVEKIDVGEAE-PRIVSGLVGFVPEKELQDLVVLV 63
Db 266 CNLKPVMNRGKSTAMVLCGSND--DK-VFVEPPKSKAGDKVFFEGFGDEAPMKOLNP 322
QY 64 CNLKPQKMRGVESOGMLLCASIEGINRQVEPLDPAGAPAEHFVAVGKGPDEELKP 123
Db 323 KKKIWEHLQHPFTND 338
QY 124 KKKYERKLQADEKISE 139

RESULT 2
ID YGJH_ECOLI STANDARD; PRT; 110 AA.
AC P42589;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 12.3 KD PROTEIN IN ILEX-EBGR INTERGENIC REGION.
GN YGJH.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
RL SUBMITTER (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: TO THE C-TERMINAL OF MET TRNA SYNTHETASES AND TO YEAST GAP1.
DR EMBL: U18997; G606012; -;
DR EMBL: AE000389; G1789455; -;
DR ECOGENE: EG12719; YGJH.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 110 AA; 12315 MW; 65065920 CRC32;

Query Match 25.7%; Score 308; DB 1; Length 110;
Best Local Similarity 50.6%; Pred. No. 1.08e-41;
Matches 41; Conservative 15; Mismatches 24; Indels 1; Gaps 1;

Db 10 ARLEMRVGYKIVYERHKNADKLYVOYDVGOKLTQYTS-LVPEYSBELMKGTVVLCN 68
QY 6 SRDIDRKGKIITVEKHPDADSLVEKIDVGEAEPRIVSGLVGFVPEKELQDLVVLV 65
Db 69 LQKRMGRTSECMILCAETD 89
QY 66 LKPQKMRGVESOGMLLCASIE 86

RESULT 3
ID SYM_THERM STANDARD; PRT; 616 AA.
AC P23395;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE) (METS).
GN THERMUS AQUATICUS (SUBSP. THERMOPHILUS).
OS PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC UNCERTAIN.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HB8;
RX MEDLINE: 91131636.
RA NUREKI O., MURAMATSU T., SUZUKI K., KOHDA D., MATSUZAWA H.,
RA OHTA T., MIYAZAWA T., YOKOYAMA S.;
RU J. BIOL. CHEM. 266:3268-3277(1991).
CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +

CC PYROPHOSPHATE + L-METHIONYL-TRNA(MET).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASES.
DR EMBL: M64273; G155136; -;
DR PIR: A39517; STYMT.
DR PROSITE: PS00178; AA_TRNA_LIGASE_1; FALSE_NEG.
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
FT SIMILAR 10 22
FT METAL 297 301 "KMSK" REGION.
FT BINDING 300 300 ATP (BY SIMILARITY).
SQ SEQUENCE 616 AA; 70638 MW; B016F5A CRC32;

Query Match 22.2%; Score 266; DB 1; Length 616;
Best Local Similarity 47.6%; Pred. No. 7.70e-33;
Matches 39; Conservative 22; Mismatches 19; Indels 2; Gaps 2;

Db 520 AKVELRAVLAEKHPNADRLVLRISGNEE-PRVSGIAWYRPEELVGKRVLVAN 578
QY 6 SRDIDRKGKIITVEKHPDADSLVEKIDVGEAEPRIVSGLVGFVPEKELQDLVVLV 65
Db 579 LKPRAKLGISOGML-AADEG 599
QY 66 LKPQKMRGVESOGMLLCASIEG 87

RESULT 4
ID SYM_METJA STANDARD; PRT; 651 AA.
AC Q58659;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE) (METS).
GN METS OR MJ1263.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96337999.
RA BUTL C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOGAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBEER R., KIRKGENS E.F., WEINSTOCK K.G., MERRICK J.M., GLADER A.,
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NOUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KLINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESSE C.R., VENTER J.C.;
RL SCIENCE 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +
PYROPHOSPHATE + L-METHIONYL-TRNA(MET).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASES.
DR EMBL: U67567; G1591898; -;
DR PROSITE: PS00178; AA_TRNA_LIGASE_1; FALSE_NEG.
DR TIGR: MJ1263; -;
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING;
KM METAL-BINDING; ZINC.
FT SIMILAR 8 20
FT METAL 320 324 "HIGH" REGION.
FT METAL 142 142 "KMSK" REGION.
FT METAL 145 145 ZINC (BY SIMILARITY).
FT METAL 154 154 ZINC (BY SIMILARITY).
FT METAL 157 157 ZINC (BY SIMILARITY).
SQ SEQUENCE 651 AA; 75558 MW; 6053C94F CRC32;

Query Match 17.6%; Score 211; DB 1; Length 651;
Best Local Similarity 40.0%; Pred. No. 1.17e-21;
Matches 30; Conservative 20; Mismatches 24; Indels 1; Gaps 1;

Db 553 KILRVRGEVVEADIRSKLKLAMVDLD-EKKQIVSGIKGYKPEDLVGKRVIVCNL 611
QY 7 RLDIRGKIITVEKHPDADSLVEKIDVGEAEPRIVSGLVGFVPEKELQDLVVLV 66

Db 612 KPAKLGVLSEGMIL 626
 QY 67 KPOKMGVSGMIL 81

RESULT 5
 ID SYM_HAEIN STANDARD; PRT; 682 AA.

AC P43828;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE) (METS).
 GN METG OR H11276.
 OS HAEMOPHILUS INFLUENZAE.
 OC PROKARYOTA: GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS; PASTURELLACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20;
 RX MEDLINE; 95350630.
 RA KERRIDGE A.R., BULL C.J., TOMB J.-F., DOUGHERTY B.A., KIRKNESS E.F., MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D., SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M., WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., CORTON M.D., UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C., RA FINE L.D., FRITCHMAN J.L., GEORHAGEN N.S.M., GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O., VENTER J.C.;
 RL SCIENCE 269:496-512(1995).
 CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR TRNA(FMET) AMINOACYLATION (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP + PYROPHOSPHATE + L-METHIONYL-TRNA(MET).
 CC -1- COFACTOR: BINDS A ZINC ION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: STRONG, TO CYSTEINYL-TRNA SYNTHETASE.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASES.
 DR EMBL; U32807; G1574731; -
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
 DR TIGR; H11276; -
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING; METAL-BINDING; ZINC.
 KM METAL-BINDING; ZINC.
 FT SIMILAR 15 25 "HIGH" REGION.
 FT SIMILAR 331 335 "KMSKS" REGION.
 FT BINDING 334 334 ATP (BY SIMILARITY).
 FT METAL 146 146 ZINC (BY SIMILARITY).
 FT METAL 149 149 ZINC (BY SIMILARITY).
 FT METAL 159 159 ZINC (BY SIMILARITY).
 FT METAL 162 162 ZINC (BY SIMILARITY).
 SQ SEQUENCE 682 AA; 77023 MM; DB6D3990 CRC32;

Query Match 16.7%; Score 200; DB 1; Length 682;
 Best Local Similarity 38.1%; Pred. No. 1.70e-19;
 Matches 32; Conservative 24; Mismatches 24; Indels 4; Gaps 4;

Db 582 AKLDMRAVAKVCEAVPSNKLRFELDLDG-HTROVSGIKAAVKNKPELEGREVIWVA 640
 QY 6 SRLDIRGKITTVEKHPADSLYEKIDVGEAEPRIVYVSGLVQVPEEELDRIVYLVC 64
 Db 641 NLAPRKMKFV-SEGMLISAGTG 663
 QY 65 NLKPOKMR-GVESOGMLCASIEG 87

RESULT 6
 ID SYM_BACST STANDARD; PRT; 649 AA.
 AC P23920;
 DT 01-MAR-1992 (REL. 21, CREATED)

DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE) (METS).
 GN METS.
 OS BACILLUS STEAROTHERMOPHILUS.
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 1518;
 RX MEDLINE; 91305115.
 RA MECHULAM Y., SCHMITT E., PANVERT M., SCHMITTER J.-M., FAYAT G., LAPADAT-TAPOLSKY M., MEINDEL T., DESSEN P., BLANQUET S.,
 RL NUCLEIC ACIDS RES. 19:3673-3681(1991).
 CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR TRNA(FMET) AMINOACYLATION.
 CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP + PYROPHOSPHATE + L-METHIONYL-TRNA(MET).
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: STRONG, TO CYSTEINYL-TRNA SYNTHETASE.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASES.
 DR EMBL; X57925; G39989; -
 DR PIR; S16682; S16682.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 KM AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 FT SIMILAR 11 23 "HIGH" REGION.
 FT BINDING 298 302 "KMSKS" REGION.
 FT BINDING 301 301 ATP (BY SIMILARITY).
 SQ SEQUENCE 649 AA; 74354 MM; 94616A5D CRC32;

Query Match 15.7%; Score 188; DB 1; Length 649;
 Best Local Similarity 37.8%; Pred. No. 3.57e-17;
 Matches 31; Conservative 23; Mismatches 27; Indels 1; Gaps 1;

Db 551 AKVLDRAVAVQPRMKRADLKLQDLG-GEKROYISGAEIKYKPELIGKAVICAN 609
 QY 6 SRLDIRGKITTVEKHPADSLYEKIDVGEAEPRIVYVSGLVQVPEEELDRIVYLVC 65
 Db 610 LKPAKLGVLSEGMILAGSGG 631
 QY 66 LKPOKMGVSGMILCASIEG 87

RESULT 7
 ID SYM_BACST STANDARD; PRT; 664 AA.
 AC P37465;
 DT 01-OCT-1994 (REL. 30, CREATED)
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE) (METS).
 GN METS.
 OS BACILLUS SUBTILIS.
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE; 96051385.
 RA OGASAWARA N., NAKAI S., YOSHIKAWA H.;
 RL DNA RES. 1:1-14(1994).
 CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR TRNA(FMET) AMINOACYLATION.
 CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP + PYROPHOSPHATE + L-METHIONYL-TRNA(MET).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: STRONG, TO CYSTEINYL-TRNA SYNTHETASE.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASES.

[illegible]

```

OY      7 RLDIRVGKILITVEKHPPDADSLYEKKIDVGAERPTVVSGL-VQFVPKEBLDRLLVVJCN 65
Db      611 LKPAKLGMSEMGML 626
           ||| | : | ||| : |
OY      66 LKPKMRGVEISOGML 81

RESULT          9
ID   SYM_ECOLI        STANDARD;             PRT;         676 AA.
AC   P00959;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DE   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)
DE   (METRS).
GN   MERG.
OS   ESCHERICHIA COLI.
OC   PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC   ENTEROBACTERIACEAE.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 85054627.
RA   DARDEL F., FAYAT G., BLANQUET S.;
RL   J. BACTERIOL. 160:1115-1122(1984).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN-K12;
RX   MEDLINE; 91080852.
RA   DARDEL F., PANVERT M., FAYAT G.;
RL   MOL. GEN. GENET. 223:121-133(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN-K12 / BHB2600;
RA   RICHTERICH P., LAKEY N., GRAYN G., JAEHN L., MINTZ L., ROBISON K.,
RA   CHURCH G.M.;
RL   SUBMITTED (OCT-1993) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN-K12 / MG1655;
RA   BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERINA N.T., GLASNER F.D.;
RL   SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [5]
RP   SEQUENCE OF 1-565 FROM N.A.
RX   MEDLINE; 83079258.
RA   BARKER D.G., EBEL J.-P., JAMES R., BRUTON C.J.;
RL   EUR. J. BIOCHEM. 127:449-457(1982).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX   MEDLINE; 82192427.
RA   ZEIMER C., RISLER J.L., BRUNIE S.;
RL   J. MOL. BIOL. 155:63-81(1982).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX   MEDLINE; 91073404.
RA   BRUNIE S., ZEIMER C., RISLER J.L.;
RL   J. MOL. BIOL. 216:411-424(1990).
RN   [8]
RP   STRUCTURE BY NMR OF 138-163.
RX   MEDLINE; 93294859.
RA   FOURMY D., DARDEL F., BLANQUET S.;
RL   J. MOL. BIOL. 231:1078-1089(1993).
RN   [9]
RP   ACTIVE SITE MAPPING.
RX   MEDLINE; 92070503.
RA   FOURMY D., MECHUAM Y., BRUNIE S., BLANQUET S., FAYAT G.;
RL   FEBS LETT. 292:259-263(1991).
RN   [11]
RP   MOTAGENESIS OF ZINC-LINGANDS.

```


RX MEDLINE: 93294858.
 RA FOUJAY D., MEINDEL T., MECHULAM Y., BLANQUET S.;
 RL J. MOL. BIOL. 231:1068-1077(1993).
 RP REVIEW.
 RX MEDLINE: 91129305.
 RA MEINDEL T., MECHULAM Y., DARDEL F., SCHMITTER J.M., HOUNTONDJI C.,
 RA BRUIE S., DESSEN P., FAYAT G., BLANQUET S.;
 RL BIOCHIMIE 72:625-632(1990).
 CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING
 REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO
 FOR THE INITIATION OF ALL RNA TRANSLATION THROUGH INITIATOR
 TRNA (FMET) AMINOACYLATION.
 CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +
 PYROPHOSPHATE + L-METHIONYL-TRNA(MET).
 CC -1- COFACTOR: BINDS A ZINC ION.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: STRONG, TO CYSTEINYL-TRNA SYNTHETASE.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASES.
 DR EMBL: K02671: G146829; -.
 DR EMBL: X55791: G42016; -.
 DR EMBL: U00007: G405895; ALT_INIT.
 DR EMBL: AE00300: G1788432; -.
 DR PIR: S1427: SYECMT.
 DR PIR: A35821: A35821.
 DR PIR: S11949: S11949.
 DR PDB: 1MEA: 31-JAN-94.
 DR PDB: 1MED: 31-JAN-94.
 DR SWISS-2DPAGE: P00959; COLI.
 DR ECODBASE: F072.0; 6TH EDITION.
 DR ECODBASE: G072.0; 6TH EDITION.
 DR ECOGENE: EG10586; METG.
 DR PROSITE: PS00178: AA-TRNA-LIGASE.I: 1
 KW AMINOACYL-TRNA SYNTHETASE, PROTEIN BIOSYNTHESIS, LIGASE; ATP-BINDING;
 KM METYL-BINDING; ZINC; 3D-STRUCTURE.
 FT INIT_MET 0
 FT SIMILAR 14 24 "HIGH" REGION.
 FT SIMILAR 332 336 "KMSKS" REGION.
 FT METAL 145 145 ZINC.
 FT METAL 148 148 ZINC.
 FT METAL 158 158 ZINC.
 FT METAL 161 161 ZINC.
 FT BINDING 335 335 ATP.
 FT MURGEN 335 335 K->Q,A,E,R: LOSS OF ACTIVITY.
 FT STRAND 141 141
 FT TURN 142 143
 FT STRAND 144 144
 FT TURN 146 148
 SQ SEQUENCE 676 AA; 76123 MW; 9D92D848 CRC32;
 Query Match 13.4%; Score 160; DB 1; Length 676;
 Best Local Similarity 33.7%; Pred. No. 6.29e-12;
 Matches 35; Conservative 26; Mismatches 37; Indels 6; Gaps 6;
 Db 576 AKVDIALVALIENAEFEVSGDRLRLTLDLG-GKKRVVFGIRAVYDPPALIGRHIMYA 634
 QY 6 SLDDIRVKGIIIVKHPDSDLVEKIDVGEAPRTIVVSGL-VQFPKRELDRLVVVL 64
 Db 635 NLAPRKRFGI-SEGVMVMAAG-PG-GKIDIFLLSPDAGARKPGHGV 675
 QY 65 NLKPKMRGVRSGMGLCASTIEGIRNOVEPLDPPAGSAPGEHV 107
 RESULT 10
 ID SYTB-SYNY3 STANDARD; PRT; 810 AA.
 AC P74296;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLALANINE--
 TRNA LIGASE BETA CHAIN) (PHERS).
 GN PHER OR SL1553.

OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
 OC PROKARYOTA; GRACILICUTES; OXYPHOTOBACTERIA;
 OC CYANOBACTERIA (BLUE-GREEN ALGAE); CHROOCOCCALES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97061201.
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARAO K.,
 RA OKUDERA S., SHIMO S., TAKEUCHI C., WADA T., MATANABE A.,
 RA YAMADA M., YASUDA M., TABATA S.;
 RL DNA RES. 3:109-136(1996).
 CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE + TRNA(PHE) = AMP +
 PYROPHOSPHATE + L-PHENYLALANYL-TRNA(PHE).
 CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
 (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 DR EMBL: D90913: G1653476; -.
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 SQ SEQUENCE 810 AA; 87887 MW; C39AC525 CRC32;
 Query Match 13.2%; Score 158; DB 1; Length 810;
 Best Local Similarity 40.7%; Pred. No. 1.45e-11;
 Matches 37; Conservative 20; Mismatches 25; Indels 9; Gaps 6;
 Db 45 VVLGRVTSREKHPNMDKLSVCYVDIGTEPRTIYCGAANVRADILVPATLGSYLPRVDL 104
 QY 10 IRVGKIIIVKHPDSDLVEKIDVGEAPRTIVVSGL-VQ--FVPEKELDRLVV-L 63
 Db 105 -KIKPAKLRGVKSGMI-CSLAELGLSKSE 133
 QY 64 CNLKPKMRGVRSGMGLCASTIEGINRQVE 93

RESULT 11
 ID SYTB-SYNY7 STANDARD; PRT; 808 AA.
 AC P74764;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLALANINE--
 TRNA LIGASE BETA CHAIN) (PHERS).
 GN SYNECHOCOCCUS SP. (STRAIN PCC 7942) (ANACYSTIS NIDULANS R2).
 OC PROKARYOTA; GRACILICUTES; OXYPHOTOBACTERIA;
 OC CYANOBACTERIA (BLUE-GREEN ALGAE); CHROOCOCCALES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA CRANENBURGH R.M., ROBINSON N.J.;
 RL J. APPL. PHYCOL. 8:81-82(1996).
 CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE + TRNA(PHE) = AMP +
 PYROPHOSPHATE + L-PHENYLALANYL-TRNA(PHE).
 CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
 (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 DR EMBL: X94345: E264731; -.
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 SQ SEQUENCE 808 AA; 89840 MW; FBAD67C5 CRC32;
 Query Match 12.9%; Score 154; DB 1; Length 808;
 Best Local Similarity 38.5%; Pred. No. 7.68e-11;
 Matches 35; Conservative 22; Mismatches 25; Indels 9; Gaps 7;
 Db 45 VVVGRLVERQHPNMDRLSVCOVEITGQAEPTIVICGASVRADIVAAVATLGSYLPCIDL 104
 QY 10 IRVGKIIIVKHPDSDLVEKIDVGEAPRTIVVSGL-VQ--FVPEKELDRLVV-L 63
 Db 105 -KIKPKLRGVKSGMI-CSLSELGLTKESE 133
 QY 64 CNLKPKMRGVRSGMGLCASTIEGINRQVE 93
 RESULT 12

ID STFB_HELPY STANDARD; PRT: 764 AA.
 AC P56145;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DE PHENYLAANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLAANYNE--
 TRNA LIGASE BETA CHAIN) (PHERS).
 GN PHE7 OR HP0402.
 OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;
 CC AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695;
 RX MEDLINE; 97394467.
 RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
 RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
 RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
 RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODER A.,
 RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
 RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
 RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATNEY L., WALLIN E.,
 RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
 RA VENTER J.C.; 339-547(1997).
 RL NATURE 388:339-547(1997).
 CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLAANYLINE + TRNA(PHE) - AMP +
 CC PYROPHOSPHATE + L-PHENYLAANYL-TRNA(PHE).
 CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 DR TIGR: HP0402;
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 SQ SEQUENCE 764 AA; 85135 MW; 93C8EB3B CRC32;
 Query Match 10.7%; Score 128; DB 1; Length 764;
 Best Local Similarity 36.8%; Pred. No. 2.51e-06;
 Matches 32; Conservative 18; Mismatches 29; Indels 8; Gaps 6;
 Db 39 IAPK-NNVVGKILEKAPKNAKNAKLVSCVDYKREVLQIVCGAKNAPAPQFVP-VALNGAL 96
 Qy 4 IPRSLDRVGLIIVKHPDADSLYVERKIDYGEAPRTVSGT-V---QVFKRELQDRL 59
 Db 97 IGSTTAKTE-LRGVESHGMI-CSSIE 121
 Qy 60 VVVLCLNKLKPKMRGVEESQGMILLASIE 86
 RESULT 13
 ID STFB_ECOLI STANDARD; PRT: 795 AA.
 AC P07395; Q59407;
 DT 01-APR-1988 (REL. 07, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DE PHENYLAANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLAANYNE--
 TRNA LIGASE BETA CHAIN) (PHERS).
 GN PHE7.
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 CC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 85261100.
 RA MECHOLMAN Y., FAYAT G., BLANQUET S.;
 RL J. BACTERIOL. 163:787-791(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MILLER H.I.;
 RL SUBMITTED (NOV-1986) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RA BLATTNER F.R., PLODKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,
 RA IEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KAKI K., KASAI H.,
 RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,
 RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,
 RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPET G., SEKI Y.,
 RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.;
 RA SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLAANYLINE + TRNA(PHE) - AMP +
 CC PYROPHOSPHATE + L-PHENYLAANYL-TRNA(PHE).
 CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 DR EMBL; V00291; G43071; -;
 DR EMBL; K02844; G146346; -;
 DR EMBL; A0000266; G1788006; -;
 DR EMBL; D90814; G1742806; -;
 DR EMBL; D90813; G1742793; -;
 DR PIR; B23099; SYCEFB.
 DR SWISS-2DPAGE; P07395; COLI.
 DR ECO2DPAGE; D094.0; 6TH EDITION.
 DR ECOGENE; EG10710; PHE7.
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 FT CONFLICT 93 93 A -> T (IN REF. 1).
 FT CONFLICT 141 142 AP -> VR (IN REF. 1).
 FT CONFLICT 186 189 PLYO -> AAGN (IN REF. 1).
 FT CONFLICT 481 481 A -> R (IN REF. 1).
 FT CONFLICT 698 698 R -> G (IN REF. 1).
 SQ SEQUENCE 795 AA; 87378 MW; 5E5C3A3C CRC32;
 Query Match 8.8%; Score 106; DB 1; Length 795;
 Best Local Similarity 29.2%; Pred. No. 7.86e-03;
 Matches 33; Conservative 22; Mismatches 51; Indels 7; Gaps 6;
 Db 36 EPVAGSHGVYGEVEVECAQHPNADKLRTKVNVGGRLLDIYCGAPNCGRLRVAVAT 95
 Qy 1 EEVIPSRLDIRVGLIIVKHPDADSLYVERKIDYGEAPRTVSGT-V---QVFKRELQDRL 59
 Db 96 GAVLPGFKIKAATLRGEPSEGM-CSEFSLGSDSHSGLIEPADAPIGTDI 147
 Qy 60 VVVLCLN-LKPKMRGVEESQGMILLASIE-GI-NROVEPLDPAGSAPGEHV 107

RESULT 14
 ID STFB_MYCPN STANDARD; PRT: 805 AA.
 AC P75563;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DE PHENYLAANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLAANYNE--
 TRNA LIGASE BETA CHAIN) (PHERS).
 GN MYCOPLASMA PNEUMONIAE.
 OC MYCOPLASMA; TENERICUTES; MYCOPLASMA; MYCOPLASMALES;
 CC MYCOPLASMATACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE; 97105885.
 RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
 RA HERRMANN R.;
 RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
 CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLAANYLINE + TRNA(PHE) - AMP +
 CC PYROPHOSPHATE + L-PHENYLAANYL-TRNA(PHE).
 CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 DR EMBL; A000006; G1673698; -;
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 SQ SEQUENCE 805 AA; 91713 MW; F4F657DC CRC32;
 Query Match 8.8%; Score 105; DB 1; Length 805;

!!SEQUENCE_LIST 1.0

(Nucleotide) FASTA of: seqla from: 1 to: 504 April 28, 1998 08:40

TO: geneseqn: * Sequences: 159,646 Symbols: 57,748,920 Word Size: 6

Sequences too short to analyze: 10 (42 symbols)

Databases searched:

Geneseq-NA, Release 30.0, Released on 26Jan1998, Formatted on 10Mar1998

Searching with both strands of the query.

Scoring matrix: GenRunData:fastadna.cmp

Constant pamfactor used

Gap creation penalty: 16 Gap extension penalty: 4

Histogram Key:

Each histogram symbol represents 787 search set sequences
Each inset symbol represents 3 search set sequences
z-scores computed from opt scores

```

z-score Obs      exp
      (=)      (*)
< 20  518      0 : *
    22  26      0 : *
    24   9      0 : *
    26   3      6 : *
    28   6      6 : *
    30  197     388 : *
    32  596    1500 : *
    34  1692   4069 : *
    36  4018   8356 : *
    38  7410  13809 : *
    40  11127 19263 : *
    42  19191 23546 : *
    44  32719 25974 : *
    46  39908 26455 : *
    48  47177 25328 : *
    50  25910 23112 : *
    52  27384 20319 : *
    54  13730 17356 : *
    56  10401 14498 : *
    58   9391 11902 : *
    60   6335  9641 : *
    62   5619  7730 : *
    64   4236  6147 : *
    66   3167  4859 : *
    68   2586  3822 : *
    70   1919  2995 : *
    72   1629  2340 : *
    74   1187  1825 : *
    76   899   1420 : *
    78   615   1104 : *
    80   590   857 : *
    82   470   656 : *
    84   312   519 : *
    86   228   402 : *
    88   170   311 : *
    90   169   241 : *
    92   90    186 : *
    94   120   144 : *
    96   58    111 : *
    98   47     86 : *
   100   36     67 : *
   102   33     52 : *
   104   26     40 : *
   106   30     31 : *
   108   28     24 : *
   110   12     19 : *

```

```

112 9 14 : * : *
114 0 11 : * :
116 0 9 : * :
118 11 7 : * :
>120 20 5 : * :

```

Results sorted and z-values calculated from opt score
1968 scores saved that exceeded 81
117963 optimizations performed
Joining threshold: 51, optimization threshold: 36, opt. width: 16

The best scores are:

```

GENESEQN:T49312      Begin: 94 End: 597      Initl Initn opt z-sc E(281635)..
! Human endothelial-monocyte activati... 2520 2520 2520 2556.0 0
GENESEQN:Q86718      Begin: 506 End: 746
! Mouse EMAP111 CDNA... New endothelia... 424 524 449 449.6 3.3e-18
GENESEQN:Q94113      Begin: 242 End: 389      Strand: -
! PML2 genomic DNA... Thrombopoietin ... 72 72 136 132.1 1.7
\\End of List

```

seqla
GENESEQN:T49312

```

ID T49312 standard; cDNA; 636 BP.
AC T49312:
DT 04-JUL-1997 (first entry)
DE Human endothelial-monocyte activating polypeptide III encoding cDNA.
KW Endothelial-monocyte activating polypeptide; EMAP 111; human;
   neoplasia; ss. . .

```

SCORES Initl: 2520 Initn: 2520 Opt: 2520 z-score: 2556.0 E(): 0
100.0% identity in 504 bp overlap

```

seqla      10      20      30
GAGAGATCATCCATCCCGGCTGGATATC
|||||
T49312      70      80      90     100     110     120
GGCCAAGGCGCTGCCAAGATTTCAGAACCAAGAGGATGATCCATCCCGGCTGGATATC
|||||
seqla      40      50      60      70      80      90
CGGTGGGGAATATCATCTGTGAGAACCAACCAATGAGAGAGCGCTTATGTATAG
CGTGGGGAATATCATCTGTGAGAACCAACCAATGAGAGCGCTTATGTATAG
130     140     150     160     170     180
seqla      100     110     120     130     140     150
AAGATTGACGTGGGGAGAGCTGAACCAAGAGCTGTGAGCGGCGCTGTACAGTTCGTG
|||||
T49312      190     200     210     220     230     240
AAGATTGACGTGGGGAGAGCTGAACCAAGAGCTGTGAGCGGCGCTGTACAGTTCGTG
190     200     210     220     230     240
seqla      160     170     180     190     200     210
CCCAAGAGAACTGCAAGACAGGCTGTGAGTGTGCTCAACCTTAAACCCAGAG
|||||
T49312      250     260     270     280     290     300
CCCAAGAGAACTGCAAGACAGGCTGTGAGTGTGCTCAACCTTAAACCCAGAG
250     260     270     280     290     300
seqla      220     230     240     250     260     270
ATGAGAGAGTGCAGTCCCAAGGCGCTGTGAGTGTGCTTATATAGAGGATTAACCGC
|||||
T49312      310     320     330     340     350     360
ATGAGAGAGTGCAGTCCCAAGGCGCTGTGAGTGTGCTTATATAGAGGATTAACCGC
310     320     330     340     350     360
seqla      280     290     300     310     320     330
CAGGTGAACCTGTGAGACCTTCGCGAGGCTCTGCTCTGTGAGAGGTTTGTGAG
|||||
T49312      370     380     390     400     410     420
CAGGTGAACCTGTGAGACCTTCGCGAGGCTCTGCTCTGTGAGAGGTTTGTGAG
370     380     390     400     410     420
seqla      340     350     360     370     380     390
GGCTATGAAAAGGGCCAAACCAAGATGAGAGCTCAAGCCCAAGAGAAGAGTTCGAGAG

```

	T49312	seq1a	T49312	seq1a	T49312	seq1a	T49312	seq1a
GENESEQN:086718	GGCTATTAAGAGGCCACAGATGAGGAGCTCAAGCCCAAGAAAGTCTTGAGAG	TTTCAGGCTGACCTTCAAATTTCTGAGGAGTGCATCGACAGTGGAAACCAACTTTC	ATGACCAAGCTGGGCTCCATTTCTCTTAATATCGCTGAAAGGGGGAAACATTAGC	ATGACCAAGCTGGGCTCCATTTCTCTTAATATCGCTGAAAGGGGGAAACATTAGC	ATGACCAAGCTGGGCTCCATTTCTCTTAATATCGCTGAAAGGGGGAAACATTAGC	ATGACCAAGCTGGGCTCCATTTCTCTTAATATCGCTGAAAGGGGGAAACATTAGC	ATGACCAAGCTGGGCTCCATTTCTCTTAATATCGCTGAAAGGGGGAAACATTAGC	ATGACCAAGCTGGGCTCCATTTCTCTTAATATCGCTGAAAGGGGGAAACATTAGC
ID	086718	standard; CDNA; 1086 BP.	460	470	480	490	500	510
AC	086718;		430	440	450	460	470	480
DT	29-SEP-1995 (first entry)		490	500	510	520	530	540
DE	Mouse EMAP11 CDNA.		430	440	450	460	470	480
KW	EMAP11; endothelial monocyte activating polypeptide II; chemotaxis; inflammation; tissue factor; tumor; cancer; therapy; metha; . . .		430	440	450	460	470	480

SCORES Init1: 424 Initn: 524 Opt: 449 z-score: 449.6 E(): 3.3e-18
65.1% Identity in 241 bp overlap

[illegible][illegible]

```

1 CPU time used:
2 Database scan: 0:10:07.9
3 Post-scan processing: 0:00:00.3
4 Total CPU time: 0:10:08.2
5 Output file: seqia.fasta
6 INNA_SEQUENCE 1.0
7 ID T49312 standard; cDNA; 636 BP.
8 AC T49312:
9 DT 04-JUL-1997 (first entry)
10 DE Human endothelial-monocyte activating polypeptide III encoding cDNA.
11 KW Endothelial-monocyte activating polypeptide; EMAP III; human;
12 KM neoplasia; ss.
13 OS Synthetic.
14 FH Key
15 FH CDS
16 FH Location/Qualifiers
17 FT CDS
18 FT /tag_ a
19 FT /product= EMAP_III
20 PN MO9640719-A1.
21 PD 19-DEC-1996.
22 PF 07-JUN-1995; U07328.
23 PR 07-JUN-1995; WO-U07328.
24 PA (HUMA-) HUMAN GENOME SCI INC.
25 PI Coleman TA, Olsen HS, Rosen CA;
26 DR WPI: 97-052219/05.
27 P-PSDB; M06596.
28 FT New isolated endothelial-monocyte activating polypeptide III - used
29 PT to develop prods. for the diagnosis and treatment of conditions
30 PT involving EMAP III, e.g. neoplasia
31 PS Claim 5: Page 37: 50pp; English.
32 CC The present sequence encodes the 168 amino acids of the novel
33 CC polypeptide endothelial-monocyte activating polypeptide III (EMAP III)
34 CC The product can be used for the diagnosis and treatment of conditions
35 CC involving abnormal expression of EMAP III. In particular, EMAP III can
36 CC be used to regress neoplasia such as in tumours and cancers.
37 CC EMAP III shows a high degree of homology to EMAP II with 60%
38 CC identity and 75% similarity over a 150 amino acid stretch.
39 CC Sequence 636 BP: 172 A; 166 C; 178 G; 120 T;
40

```

```

T49312 Length: 636 April 28, 1998 12:50 Type: N Check: 4493 ..
1 TACCCCTGCC CTGAAAAAC TGGCCAGCCG TGGCTACCCA GATCCCTCAA
51 AGCAGAACCC AATGGCCAAA GGCCTGCCAA GAATTCAGAA CCAAGAGGAG
101 TCATCCCATC CCGGCTGGAT ATCCGTGTGG GGAATATCAT CACTGTGGAG
151 AAGACCCAG ATGCAGACAG CCTGTATGTA GAGAAGATTG ACGTGGGGGA
201 AGCTGAACCA CGGACTGTGG TGAGCGGCTT GGTACAGTTC GTGCCCAAGG
251 AGGAAGTCCA GAGACGGCTG GTAGTGGTGC TGTGCAACCT GAAACCCGAG
301 AAGATGAGAG GAGTGGAGTC CCAAGGCATG CTTGTGTGTG CTTGTATGTA
351 AGGATAAAC CGCCAGTTGG AACCTGTGGA CCTCCGGGA GGCCTGTGTC
401 CTGTGAGACA CGTGTGTGTG AAGGGCTATG AAAAGGGCCA ACCAGATGAG
451 GAGCTCAAGC CCAAGAGAA AGCTTCGAG AAGTGCAGG CTGACTTCAA
501 AATTTCTGAG GAGTGCATCG CACAGTGAA GCAAAACCAAC TTCAATGACCA
551 AGCTGGGCTC CATTTCTGT ATATCGCTGA AAGGGGGGNA CATTAGCTAG
601 CCAGCCACAG ATCTTCCCC CTCTCTCCAC CACTGA

!!NA_SEQUENCE 1.0
ID 086718 standard; cDNA; 1086 BP.
AC 086718;
DT 29-SEP-1995 (first entry)
DE Mouse EMAP11 cDNA.
KW EMAP11; endothelial monocyte activating polypeptide II; chemotaxis;
KW inflammation; tissue factor; tumor; cancer; therapy; metha;
KW sarcoma; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 64..996
FT /*tag- a
PN W09509180-A.
PD 06-APR-1995.
PE 29-SEP-1994; U11085.
PR 29-SEP-1993; US-129456.
PA (UNC) UNIV COLUMBIA NEW YORK.
PI Claus M, Rao J, Kayton M, Liputti SK, Stern DM;
DR WPI: 95-147389/19.
DR P-PSDB: R72577.
PT New endothelial monocyte activating polypeptide II - induces
PT chemotaxis, inflammation and tissue factor, useful for treating
PT tumours, also related antibodies, DNA and active fragments
PS Disclousure: Fig.4; 180pp; English.
CC A mouse metha sarcoma cDNA library was screened with a probe
CC based on the N-terminal sequence of mouse EMAP11. Overlapping
CC clones were combined to obtain a contiguous full-length sequence
CC (given in 086718) encoding a 33 kDa protein (R72577).
CC Recombinant EMAP11 was expressed in E. coli.
CC Sequence 1086 BP; 346 A; 212 C; 297 G; 231 T;
SO

086718 Length: 1086 April 28, 1998 12:50 Type: N Check: 3316 ..
1 GAGGCTGTC AAGAGCTGCG GTTGGGTCAC CGCTTCATCT TTTCTGCGC
51 ATTCTGGGGA AAGATGCCAA CGAATGATGC TGTCTGANG AGGCTGAGC
101 AGAAGGCTGC AGAGCGCGAT CAGATCATCG AATATCTCAA GCAGCAGGT
151 GCTCTCTCTA AGGAGAAAGC AATTTGCAG GCAACAATA GAGAAGAAA
201 GAACTTGA GTTGAATG CTAACCTGAA AAAAGAAATA GAAGAGCTAA
251 AGCAGAGCT GATTCTGCA GAATTCATA ACGAGTGA GCAAGTGGT

```

```

301 GTTCGATTGA GACTCCACT GCAGACGAAC TGTACTGCTT CTGAAGTGT
351 GGTGAGTCT CCATCAGTAG CAACCAACCC CTCTCTCTCT ACAAAAAGC
401 AGATCAAGC GGGAGAGAA AAGAGGTGA AAGAGAGAC TGAAGAGAA
451 GGAGAGAAA AGGAGAGCA GCAGTGGCA GCAGCAAGTA CTGACTCCA
501 GCTATCCAC GCATCGCTC TGGATCTTG AATTGGTGT ATTCTACTG
551 CCAAGACCA CCTGATGCA GATTCAGTGT ATGTGAGAGA AGTAGATGTG
601 GGAGAGAGAG CCCGCGCAC GGTGTGACG GGGCTGTGTA ATCATGTTCC
651 TCTAGAACG ATGCAATC GTATGTGT TTTACTCTGT AATGTAGAC
701 CTGCAAGAT GCGGGAGTT CTGTCTCAAG CCATGGTAT GTGTGCCAGT
751 TCACCAGAG AAGTGAGAT TGTGGCCCT CCACAGGGT CCGTCTGTG
801 GGACGAATT ACTTTGATG CTTTCTCTG AGAGCTTAC AAGAGCTAA
851 ACCCTAAGAA GAAGATCTG GAGCAGATCC AGCTGACCT GCACACCAAT
901 GGTAGTGTG TGGCACATA CAAGGAGCT CCTTTGAG TGAAGGGAA
951 GGGATTGTC AGAGCCAAA CCATGGCCA TAGTGAATT AATATGTC
1001 TCTGTACTG AAGACATG GCGAAACTT AATAACAATA AAGAGAGTG
1051 TGTATATC TTACATATA AAAAAAAAA AAAAAA

!!NA_SEQUENCE 1.0
ID 094113 standard; DNA; 1014 BP.
AC 094113;
DT 27-FEB-1996 (first entry)
DE PML2 genomic DNA.
DE PML2 genomic DNA.
KW Human; thrombopoietin; TPO; mpl ligand; hml; fragment polypeptide;
KW megakaryocytolethic cytokine receptor; thrombopoietic signal;
KW EPO-domain fragment; erythropoietin; hepo; haematopoietic cell;
KW megakaryocyte; thrombocytopenia; myeloproliferative disease;
KW inflammatory thrombocytosis; iron deficiency; Epo; platelet;
KW red blood cell; progenitor; hml-2; ss.
OS Sus scrofa.
FH Key Location/Qualifiers
FT CDS 1..987
FT /*tag- a
FT /product- PML2
PN GB2285446-A.
PD 12-JUL-1995.
PE 21-DEC-1994; 025831.
PR 03-JAN-1994; US-176553.
PR 21-JAN-1994; US-185607.
PR 15-FEB-1994; US-196689.
PR 04-APR-1994; US-223263.
PR 25-MAY-1994; US-249376.
PR 02-DEC-1994; US-348658.
PR 02-DEC-1994; US-348657.
PA (GETH ) GENENTECH INC.
PI Eaton DL, de Sauvage FJ;
DR WPI: 95-234018/31.
DR P-PSDB: R76170.
PT Thrombopoietin polypeptide, ligand for mpl cytokine receptor -
PT useful for treating thrombocytopenia and related diseases
PS Example 13; Fig 21; 192pp; English.
CC This sequence represents a genomic clone encoding an isoform of porcine
CC thrombopoietin (TPO), also known as mpl ligand (PML), pml2. This
CC sequence was isolated by RACE PCR using the primer sequences given in
CC 094140-43 and 094128. The genomic sequence was isolated from a porcine
CC genomic library in EMBL3 by screening with pR45. Two forms of pml
CC isolated in this manner. The first was a full length clone (see also
CC 094112) and the second encoded a protein with a four amino acid deletion

```

CC (pML2). Comparison of the sequences of pML and pML2 shows that the
CC latter form is identical except for the deletion of the tetrapeptide QLRP
CC corresponding to residues 111-114. The four amino acid deletions
CC observed in murine, human and porcine ML cDNA occur at precisely the
CC same position in the predicted proteins (see also Q94110). pML binds to
CC mPL which is a megakaryocytopoietic cytokine receptor which may transduce
CC a thrombopoietic signal.
SQ Sequence 1014 BP: 216 A; 365 C; 220 G; 213 T;

094113 Length: 1014 April 28, 1998 12:50 Type: N Check: 5829 ..

```
1 AGCCCGGCTC CTCCTGCTG TGACCCCGCA CTCCTAATA AACTGCTTCG
51 TGACTCCCAT GTCTTCACG GCAGACTGAG CCACTGCCA GACATTAAC
101 CTTGTCCAC ACCTGTCCTG CTGCCTGCTG TGGACTTCAC CTGGGAGAA
151 TGGAAATCCC AGACGGAGCA GACAAAGCA CAGATGTCC TGGAGCCAC
201 AACCTTCTG CTGGAGCAG TGATGACAGC ACGGGACAA GTGGAGCCCC
251 CTGGCTCTC ATCCCTGCTG GTGCAGCTT CTGGACAGT TCGCTCTCTC
301 CTGGGGCCC TGCAGGACT CCTTGAATG CAGGGAAGA CCAAGCTCA
351 CAAGGATCCC AGTGCATCT TCCTGACTT CCAACAACCTG CTCGAGGAA
401 AGGTGGCTT CCGCTGCTT GTAGTGGGGC CCTCCTCTG TGCCAAGAG
451 GCCCCACCCG CCATAGCTGT CCCGAGCAGC ACCTTCCAT TCCACACACT
501 GAACAAGCTC CCAACAGCA CCTGTGATT GTTGAGACA AACTCCAGTA
551 TCTAGCCAG AACTACTGC TCTGATTTC TCAAGAGCT GCAGGCATTG
601 AGAGCCAGA TTCCTGCTCT GCTGACCAA ACCTCAGGT CCCTAGACCA
651 AATCCCTGGA CACAGNATG GGACACACGG ACCCTGAGT GGAATTCATG
701 GACTCTTTC TGGACCCCAA CCCGGGGGCC TCGAGCTCC AGACATTCTT
751 CCAGCAACTT CAGGATGGG CTCCGGGCA ACCTACCTCC AGCCTGAGAG
801 GTCTCTTCC CCAGCTCAC CTCTCTCTG ACGATACACT CTCTCTCTC
851 CTCACCCAC CTCGCCCTCC CCCACAGTCC AGCTCCAGCC TCTGCTTCTT
901 GACCCCTCTG CGATCACACC CAACTCTACC AGTCTCTTC TATTTCAGC
951 TCACCTCAT TTCAGAACCC TGTTCAGAG AGAGTAAGT GCTCAGACCC
1001 TGCCAACCTC AGCA
```


!SEQUENCE_LIST 1.0

(Peptide) FASTA of: pep2 from: 1 to: 168 April 27, 1998 08:09

TO: geneseqp:* Sequences: 111,726 Symbols: 13,889,129 Word Size: 2

Databases searched:

Geneseq-AA, Release 30.0, Released on 26Jan1998, Formatted on 10Mar1998

Scoring matrix: GenRunData:blotsum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 361 search set sequences
Each inset symbol represents 2 search set sequences
z-scores computed from opt scores

```

z-score obs exp
(=) (=)
< 20 21641 0 : *
22 13 0 : *
24 60 0 : *
26 86 2 : *
28 331 20 : *
30 353 119 : *
32 1069 461 : *
34 1464 1250 : *
36 2212 2567 : *
38 3343 4242 : *
40 3925 5917 : *
42 6273 7232 : *
44 8758 7978 : *
46 8775 8126 : *
48 8937 7779 : *
50 9938 7099 : *
52 6337 6241 : *
54 5496 5331 : *
56 4240 4453 : *
58 3138 3656 : *
60 2486 2961 : *
62 1770 2374 : *
64 1576 1888 : *
66 1265 1492 : *
68 1183 1174 : *
70 824 920 : *
72 722 719 : *
74 443 560 : *
76 418 436 : *
78 377 339 : *
80 283 263 : *
82 173 201 : *
84 166 160 : *
86 136 123 : *
88 88 96 : *
90 58 74 : *
92 64 57 : *
94 33 44 : *
96 41 26 : *
98 34 26 : *
100 48 20 : *
102 43 16 : *
104 39 12 : *
106 32 9 : *
108 20 7 : *
110 25 6 : *
112 23 4 : *
114 21 3 : *

```

```

116 28 3 : *
118 11 2 : *
>120 34 2 : *

```

Results sorted and z-values calculated from opt score
1645 scores saved that exceeded 78
33105 optimizations performed
Joining threshold: 36, optimization threshold: 24, opt. width: 16

The best scores are:

```

GENSEQP:W06596 Begin: 1 End: 168      init1 intn opt z-sc E(86504)..
i Human endothelial-monocyte activati... 1111 1111 1111 1640.3 0
GENSEQP:W14561 Begin: 7 End: 162      401 546 585 865.4 0
i Endothelial monocyte activating pol... 398 543 582 855.4 1.2e-40
GENSEQP:R72577 Begin: 151 End: 306     101 101 176 250.6 3.3e-07
i Mouse EMAP11. New endothelial mon... 136 136 157 226.2 7.4e-06
GENSEQP:W20051 Begin: 560 End: 654     125 147 160 225.3 8.4e-06
i Methionyl-CRNA synthetase from Stap... 136 136 157 226.2 7.4e-06
GENSEQP:W20570 Begin: 338 End: 412     136 136 157 226.2 7.4e-06
i H. pylori cytoplasmic protein 78543... 125 147 160 225.3 8.4e-06
GENSEQP:W21896 Begin: 33 End: 139      136 136 157 226.2 7.4e-06
i Phenylalanyl tRNA synthetase beta s... 136 136 157 226.2 7.4e-06
GENSEQP:W20999 Begin: 564 End: 638     136 136 157 226.2 7.4e-06
i H. pylori cytoplasmic protein, hpa... 136 136 157 226.2 7.4e-06
GENSEQP:R13875 Begin: 72 End: 199      136 136 157 226.2 7.4e-06
i Murine Cytotoxic Cell Protease-2... 136 136 157 226.2 7.4e-06
GENSEQP:R42379 Begin: 16 End: 98       136 136 157 226.2 7.4e-06
i Haemin-binding protein produced as ... 136 136 157 226.2 7.4e-06
GENSEQP:R42374 Begin: 18 End: 100      136 136 157 226.2 7.4e-06
i Haemin-binding protein from ORF5 of... 136 136 157 226.2 7.4e-06
GENSEQP:R69760 Begin: 30 End: 110      136 136 157 226.2 7.4e-06
i PMON5973 (Met-15-125HIL-3) having a... 136 136 157 226.2 7.4e-06
GENSEQP:R78335 Begin: 29 End: 109      136 136 157 226.2 7.4e-06
i Human IL-3 mutant for multi-lineage... 136 136 157 226.2 7.4e-06
GENSEQP:R56050 Begin: 29 End: 109      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon133... 136 136 157 226.2 7.4e-06
GENSEQP:R80381 Begin: 42 End: 122      136 136 157 226.2 7.4e-06
i Interleukin-3 variant... Recombinan... 136 136 157 226.2 7.4e-06
GENSEQP:R80383 Begin: 43 End: 123      136 136 157 226.2 7.4e-06
i Interleukin-3 variant... Recombinan... 136 136 157 226.2 7.4e-06
GENSEQP:R56095 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon134... 136 136 157 226.2 7.4e-06
GENSEQP:R56105 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon133... 136 136 157 226.2 7.4e-06
GENSEQP:R56098 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon134... 136 136 157 226.2 7.4e-06
GENSEQP:R56099 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon134... 136 136 157 226.2 7.4e-06
GENSEQP:W13009 Begin: 398 End: 479      136 136 157 226.2 7.4e-06
i Segment of desmosomal cadherin, des... 136 136 157 226.2 7.4e-06
GENSEQP:R56096 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon134... 136 136 157 226.2 7.4e-06
GENSEQP:R56090 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon134... 136 136 157 226.2 7.4e-06
GENSEQP:R56091 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon134... 136 136 157 226.2 7.4e-06
GENSEQP:R78352 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human IL-3 mutant for multi-lineage... 136 136 157 226.2 7.4e-06
GENSEQP:R56137 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon134... 136 136 157 226.2 7.4e-06
GENSEQP:R78368 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human IL-3 mutant for multi-lineage... 136 136 157 226.2 7.4e-06
GENSEQP:R56120 Begin: 28 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon133... 136 136 157 226.2 7.4e-06
GENSEQP:R56067 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon132... 136 136 157 226.2 7.4e-06
GENSEQP:R55083 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon134... 136 136 157 226.2 7.4e-06
GENSEQP:R56136 Begin: 31 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon134... 136 136 157 226.2 7.4e-06
GENSEQP:R56115 Begin: 28 End: 110      136 136 157 226.2 7.4e-06
i Human Interleukin-3 mutein (pmon133... 136 136 157 226.2 7.4e-06

```

```

GENESQ:R10727      Begin: 55      End: 135
I (Glu108Asp109Glu110)IL-3. New int...
GENESQ:R14839      Begin: 115      End: 192
I Protein deduced from human Glioblas...
GENESQ:R78367      Begin: 43      End: 122
I Human IL-3 mutant for multi-lineage...
GENESQ:R56082      Begin: 44      End: 123
I Human Interleukin-3 mutleln (pMon13...
GENESQ:R72583      Begin: 2      End: 15
I EMAP11 active peptide. New endothe...
GENESQ:R56093      Begin: 31      End: 110
I Human Interleukin-3 mutleln (pMon13...
GENESQ:R56068      Begin: 31      End: 110
I Human Interleukin-3 mutleln (pMon13...
GENESQ:R56092      Begin: 31      End: 110
I Human Interleukin-3 mutleln (pMon13...
GENESQ:R56101      Begin: 31      End: 110
I Human Interleukin-3 mutleln (pMon13...
GENESQ:R78353      Begin: 31      End: 110
I Human IL-3 mutant for multi-lineage...
GENESQ:R56100      Begin: 31      End: 110
I Human Interleukin-3 mutleln (pMon13...
GENESQ:R56104      Begin: 31      End: 110
I Human Interleukin-3 mutleln (pMon13...
GENESQ:R56094      Begin: 31      End: 110
I Human Interleukin-3 mutleln (pMon13...
GENESQ:R22820      Begin: 44      End: 124
I K64, P83, M88, K113, E122, M127, E1...
GENESQ:R56139      Begin: 31      End: 110
I Human Interleukin-3 mutleln (pMon13...
GENESQ:W13514      Begin: 6      End: 97
I Interleukin-3 receptor agonist pMON...
GENESQ:R56141      Begin: 31      End: 88
I Human Interleukin-3 mutleln (pMon13...
GENESQ:R36739      Begin: 15      End: 60
I Human bone morphogenetic protein BM...
\\End of List

```

```

pep2
GENESQ:W06596
ID W06596 standard; Protein: 168 AA.
AC W06596;
DT 04-JUL-1997 (first entry)
DE Human endothelial-monocyte activating polypeptide III.
KW Endothelial-monocyte activating polypeptide; EMAP III; human;
neoplasta. . .

```

```

SCORES      Init1: 1111      Initn: 1111      Opt: 1111      z-score: 1640.3 E(): 0
Smith-Waterman score: 1111; 100.0% identity in 168 aa overlap

```

```

pep2      10      20      30      40      50      60
EEVIPSRLDIRVVKIITVEKHPDADSLYVEKIDVGEAEPTVVSGLVQFVKPELDRLY
|||||
W06596      10      20      30      40      50      60
EEVIPSRLDIRVVKIITVEKHPDADSLYVEKIDVGEAEPTVVSGLVQFVKPELDRLY
|||||
pep2      70      80      90      100      110      120
VVLNLKPKQMRGVESQGMGLCASIEGINRQVEPLDPAPASAPGEHVFVYKGEKGPDEE
|||||
W06596      70      80      90      100      110      120
VVLNLKPKQMRGVESQGMGLCASIEGINRQVEPLDPAPASAPGEHVFVYKGEKGPDEE
|||||

```

```

pep2      130      140      150      160
LKPKKVFEKLDQADFKISEECIAQWKOTNEMTKGISCSKLGKNGIS
|||||
W06596      130      140      150      160
LKPKKVFEKLDQADFKISEECIAQWKOTNEMTKGISCSKLGKNGIS
|||||
pep2      130      140      150      160
LKPKKVFEKLDQADFKISEECIAQWKOTNEMTKGISCSKLGKNGIS
|||||
W06596      130      140      150      160
LKPKKVFEKLDQADFKISEECIAQWKOTNEMTKGISCSKLGKNGIS
|||||

```

```

ID W14561 standard; Protein: 166 AA.
AC W14561;
DT 01-DEC-1997 (first entry)
DE Endothelial monocyte activating polypeptide II.
KW EMAP-II; inhibitor; endothelial cell formation; blood vessel;
retinopathy; tumour; subcutaneous; intraperitoneal; intravenous; . .

```

```

SCORES      Init1: 401      Initn: 546      Opt: 585      z-score: 865.4 E(): 0
Smith-Waterman score: 585; 54.0% identity in 161 aa overlap

```

```

pep2      10      20      30      40      50      59
EEVIPSRLDIRVVKIITVEKHPDADSLYVEKIDVGEAEPTVVSGLVQFVKPELDRLY
|||||
W14561      10      20      30      40      50      59
EEVIPSRLDIRVVKIITVEKHPDADSLYVEKIDVGEAEPTVVSGLVQFVKPELDRLY
|||||

```

```

pep2      60      70      80      90      100      110      119
VVLNLKPKQMRGVESQGMGLCASIEGINRQVEPLDPAPASAPGEHVFVYKGEKGPDEE
|||||
W14561      60      70      80      90      100      110      119
VVLNLKPKQMRGVESQGMGLCASIEGINRQVEPLDPAPASAPGEHVFVYKGEKGPDEE
|||||

```

```

pep2      120      130      140      150      160
ELKPKKVFEEKLDQADFKISEECIAQWKOTNEMTKGISCSKLGKNGIS
|||||
W14561      120      130      140      150      160
ELKPKKVFEEKLDQADFKISEECIAQWKOTNEMTKGISCSKLGKNGIS
|||||

```

```

pep2
GENESQ:R72577
ID R72577 standard; Protein: 310 AA.
AC R72577;
DT 29-SEP-1995 (first entry)
DE Mouse EMAP11.
KW EMAP11; endothelial monocyte activating polypeptide II; chemotaxis;
inflammation; tissue factor; tumor; cancer; therapy; metha; . .

```

```

SCORES      Init1: 398      Initn: 543      Opt: 582      z-score: 855.4 E(): 1.2e-40
Smith-Waterman score: 582; 54.0% identity in 161 aa overlap

```

```

pep2      10      20      30
EEVIPSRLDIRVVKIITVEKHPDADSLYVEKIDV
|||||
R72577      10      20      30
EEVIPSRLDIRVVKIITVEKHPDADSLYVEKIDV
|||||

```

```

pep2      40      50      60      70      80      90
EAEPRTVSGLVQFVKPELDRLYVVLNLKPKQMRGVESQGMGLCASIEGINRQVEPL
|||||
R72577      40      50      60      70      80      90
EAEPRTVSGLVQFVKPELDRLYVVLNLKPKQMRGVESQGMGLCASIEGINRQVEPL
|||||

```

```

pep2      100      110      120      130      140      150
DEPAGSAPGEHVFVYKGEKGPDEELKPKKVFEEKLDQADFKISEECIAQWKOTNEMTKG
|||||
R72577      100      110      120      130      140      150
DEPAGSAPGEHVFVYKGEKGPDEELKPKKVFEEKLDQADFKISEECIAQWKOTNEMTKG
|||||

```

```

pep2      160      170      180      190      200      210      220      230
SISCSKLGKNGIS
:|:|:|:|:|
R72577      160      170      180      190      200      210      220      230
SISCSKLGKNGIS
:|:|:|:|:|

```

```

pep2      240      250      260      270      280      290
APPNQSVPEDRITFDLFP--GEPPDKELNPKKIMEDQIDPLHNAECVATYKGAPEVK-G
|||||
R72577      240      250      260      270      280      290
APPNQSVPEDRITFDLFP--GEPPDKELNPKKIMEDQIDPLHNAECVATYKGAPEVK-G
|||||

```

```

pep2      300      310
KGVCAQTMANSGIK
:|:|:|:|
R72577      300      310
KGVCAQTMANSGIK
:|:|:|:|

```

```

pep2
GENESQ:W20051
ID W20051 standard; Protein: 657 AA.
AC W20051;
DT 10-SEP-1997 (first entry)

```

DE Methionyl-tRNA synthetase from *Staph. aureus*.
KW tRNA synthetase; *Bacillus subtilis*; mets; Immunological response;
KW antibody; bacterial infection; adherence; damaged tissue; . . .

SCORES Int1: 101 Intn: 101 Opt: 176 z-score: 250.6 E(): 3.3e-07
Smith-Waterman score: 176; 32.3% identity in 96 aa overlap

pep2 10 20 30
EEVIPSRLDIRVGKIITVEKHPDADSLYEKIDVGE

pep2
AEERTVSGIYGVCPKEELQDLVVYLNCNRKPKRGVTSQGMLTCSIGSINRQYE-PL
:
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
SEQRQLVSIAKRYETPDDLTGKRVAVNTNKRPAKLGMGRKSGMILLSEKKGVLIVSLPS
W20051
590 600 610 620 630 640

pep2 100 110 120 130 140 150
DPPASPGEHYFVNGYKGPDEELKPKKVFETLQADFKISECIAQNKQNFMTKLQ
| | :
W20051 AIPNGAVIK
650

pep2
GENESEQP:W20570

ID	W20570 standard; Protein; 436 AA.
AC	W20570;
DT	17-JUL-1997 (first entry)
DE	H. pylori cytoplasmic protein 785437.aa.
KW	Cytoplasmic; vacuole; prevention; treatment; infection; identification; . . .
	binding compound; bacterium; life cycle; activator; bacteria; inhibitor; . .

SCORES InIt1: 136 InItN: 136 Opt: 157 z-score: 226.2 E(): 7.4e-06
Smith-Waterman score: 157, 34.7% identity in 75 aa overlap

```

pep2      10 20 30
W20570    EEALAEAKEREKAPPTQENYISDEBKKYKEIVNSIGAKQGLIEKSNLIRKNDLGG
          ::::| | :: |::| |::| |

```

pep2
AEPRTVSGLVQCPKPEELQADRLVYLCLNKAKQKRGVSGQMILCASIEGINKQVEPLT
| : : : : | : : : : | : : : : | : : : : |
W20570 NRLRQIIISGIALDYEESLVGQMCVNAIKRAKIMGMSGMLAVRDNDNALISPTPR
370 380 390 400 420

pep2 100 110 120 130 140 150
P P A G S A N G E H V F V K G E K G Q P D E L K P K K Y F E K I Q A P F I S E C I A Q W Q T N F M R K L G S
W20570 430
E K I N G S L I S

pep2
GENESEQP:W21896

ID	W21896	standard; Protein; 800 AA.
AC	W21896;	
DT	10-SEP-1997	(first entry)
DE	phenylalanyl tRNA synthetase beta subunit.	
KW	tRNA synthetase; Bacillus subtilis; immunological response; antibody; bacterial infection; adherence; damaged tissue; wound healing; skin; . .	

SCORES Init1: 125 Initn: 147 Opt: 160 z-score: 225.3 E(): 8.4e-06
 Smith-Waterman score: 161, 31.7% identity in 142 aa overlap

10 20

pep2
W21896

EEVIPSRLDIR--VGKIITVEKHPDADSL
...| ||: || : |||||:|
ISNWLKEYIVTIDDSVSDLAERITPTGIEVDLIDYTDIDIKNLVGVFKSSEKHPDADKL

pep2
W21896

YVEKIDVGEAEPVTVGIV-----QFVKPEELDRILAVVLNLKPQKMGVSEGGMLLCA
|:::|||||::||::||::||::||::||
NVCVDIDGEDEPQIYICGAPNTDAGQVIYAKVCGSRLEGGI-KIKRAKLNGESESEMI-C-
70 80 90 100 110 120

pep2 90 100 110 120 130
ASTIE-GINRVEYDLPDPA-----SAGEHVEYVKGEGKQPOE-ILTRKKVFEKLAQ

w21696 130 140 150 160 170 180
SLPEIGISSNYIKFSESSIVYFSESQYPCGIDALQALYLDQVEFLLTRRDALSMIG

pep2 DFKISECIAQWKQTNFMTKLGISCSKSLKGNIS

W21896 TAYEVALINTKMTKPEETSNELELSANDELTVTIENEDKVPYYSARVVDVTIEPSPIMW
190 200 210 220 230 240

pep2
GENESEQP:W20999

ID	W20999	standard; protein; 662 AA.
AC	W20999;	
DT	21-JUL-1997	(first entry)
DE	H. pylori cytoplasmic protein, hp4el452orf1.	
KW	Cytoplasmic; vaccine; prevention; treatment; infection; identification; binding compound; bacterium; life cycle; activator; bacteria; inhibitor; . . .	

SCORES Initl: 136 Initn: 136 Opt: 157 z-score: 222.5 E(): 1.2e-05
Smith-Waterman score: 157; 34.7% identity in 75 aa overlap

pep2
W20999

	10	20	30
	EVIPSRDIIIVGRIIVKHPDADSLYEKEDIGE		
	::: ::: 1-1-		
EEALAEAKKKKKKAPPTQENITIEDFKVEIKGLIKQAQISNNKLKLTDLSE			
540	550	560	570
		580	590

pep2
W20999

	40	50	60	70	80	90
AEPPTVSGLVGCVGFPEELQDRLVAVLNLKPKMGREGSOGMLCASIGIRNVEPLD						
: : : : :	:	:	:	:	:	:
NRLRQITSSIALDYEEESLVGMGVYANLFPALIMGEMSGMILAYRDNDNLALSPPRR						
600	610	620	630	640	650	

pep2 100 110 120 130 140 150
P P A G S A P G E H V F K V G E K G Q P D E L L K P K K V F E K L Q A D F K I S E C I A Q W K T N F M K I L S

w20999 660
E K I A G S L I S

pep2
GENESEQP:R13875

ID R13875 standard; Protein; 248 AA.
 AC R13875;
 DT 11-OCT-1991 (first entry)
 DE Murine Cytotoxic Cell Protease-2.
 KM mouse; CCP2 inhibitor; cytotoxic T-lymphocytes; ss
 OS Mus musculus. . . .

```
SCORES      Initl:      53      Initn:      53      Opt:      101      z-score: 148.8      E():      0.15
Smith-Waterman score: 103;      25.7%      identity in 148 aa overlap
```

pep2

	20	30	40	50	60	70
RVGRITVEKHPDADSLYEKIDVGAEPRIVVSGLVFYFKEIDLRLVLTCLNPKPKR						
KVGGKKMFCGCGELVRDKFVLTAHHCKGSSMTVTGAHNIAAKETQOIIIPVAALPHDPY			:		:	:

```

pep2      50      60      70      80      90     100
      80      90      100      110      120
MKGESGMLCASISGINRQ--VEPLDPAGSA---PGEHVFVKYEGKQPEELKP--
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
NPDGRNDITMLKLKLVNNAKTRAVPLNDRNAHYKPGECTYAGAGKVTGDEGPKTL
110      120      130      140      150      160

pep2      130      140      150      160
-----KKKVE-KLQADF-KISEECI--AQMKTNFTKLGS-ISCKSLKGNIS
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
HEVKLVQGDQVCESEFOSSYNPAEICYGDSKIKAGSEEDSGGVCKRAAGSYSG
170      180      190      200      210      220

R13875      QTDGSAQVETRYLSEFVSWIKTKTKHS
230      240

pep2
GENESQ:R42379
ID      R42379 standard; Protein: 176 AA.
AC      R42379;
DE      19-APR-1994 (first entry)
KW      Haemlin-binding protein produced as fusion protein in pGCH5.
        Haemophilus somnus; immunogenic; haemolysin; lppB; lppC;
        thromboembolic meningoencephalitis; septicemia; arthritis; . . .

SCORES      Initl: 36 Initn: 36 Opt: 89 z-score: 134.1 E(): 1
Smith-Waterman score: 89; 23.9% identity in 92 aa overlap

pep2      30      40      50      60      70      80      89
      90      100      110      120      130      140
ROVEPLDPAPGAPGEHVFVKYEGKQPD-EELKPKKKYFEKQA-----DFKISECIA
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
R--DPYHCPAD-----VLTVGISTEANGKNIDPKKRISDKRIARMAIDLRLAECVY
50      60      70      80      90

R42379      150      160
      100      110      120      130      140      150
OMKQTNFTKLGSISCKSLKGNIS
::
RGNGKNLPOGAFDAVSTFNVGCGKMKSTLFKQANGFTPOLCHQPERMTYAGKTL
100      110      120      130      140      150

pep2
GENESQ:R42374
ID      R42374 standard; Protein: 178 AA.
AC      R42374;
DE      19-APR-1994 (first entry)
KW      Haemlin-binding protein from ORF5 of plasmid pRAP501.
        Haemophilus somnus; immunogenic; haemolysin; lppB; lppC;
        thromboembolic meningoencephalitis; septicemia; arthritis; . . .

SCORES      Initl: 36 Initn: 36 Opt: 89 z-score: 134.0 E(): 1
Smith-Waterman score: 89; 23.9% identity in 92 aa overlap

pep2      30      40      50      60      70      80      89
      90      100      110      120      130      140
EKIDVGEAPRRTVSGLVQFVKPEELQDRLVVLVLCNLKPKMGVDSQGMILCASIEGIN
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
MNFTRKMGTAICSVAAIILVYKANHQLR-ISOQGLDLIGNVEGCR
10      20      30      40

R42374      90      100      110      120      130      140
      150      160
ROVEPLDPAPGAPGEHVFVKYEGKQPD-EELKPKKKYFEKQA-----DFKISECIA
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
R--DPYHCPAD-----VLTVGISTEANGKNIDPKKRISDKRIARMAIDLRLAECVY
10      20      30      40

```

```

pep2      50      60      70      80      90
      150      160
OMKQTNFTKLGSISCKSLKGNIS
::
RGNGKNLPOGAFDAVSTFNVGCGKMKSTLFKQANGFTPOLCHQPERMTYAGKTL
100      110      120      130      140      150

pep2
GENESQ:R69760
ID      R69760 standard; Protein: 112 AA.
AC      R69760;
DE      05-OCT-1995 (first entry)
        PMON5973 (Met-15-125HIL-3) having an M7 substitution.
        Hybrid; human; murine; interleukin 3; IL-3; deletion; substitution;
        biological activity; side effect profile; leukotriene; release; . . .

SCORES      Initl: 44 Initn: 44 Opt: 81 z-score: 126.4 E(): 2.7
Smith-Waterman score: 81; 28.4% identity in 81 aa overlap

pep2      30      40      50      60      70      80
      90      100      110      120      130      140
-IEGINROYEPLDPAPGAPGEH-VFVKYEGKQPEELKPKKKYFEKQAQDFKISEECI
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
AIESILKNLPLCLPLATAPTRIHIKGDWNEFRKRTFYKLTLENQAQO
60      70      80      90      100      110

R69760      150      160
      10      20      30      40      50
MNCNMIDEITTHLKQPLPLDFENNLNGEDODILMNNLRPNLEAFNRSGEVDNMS
10      20      30      40      50

pep2      90      100      110      120      130      140
      150      160
AQMKTNFTKLGSISCKSLKGNIS

pep2
GENESQ:R78335
ID      R78335 standard; Protein: 111 AA.
AC      R78335;
DE      27-JUN-1996 (first entry)
        Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
        Interleukin-3; mutain; colony stimulating factor; CSF; chemotherapy;
        radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody; . . .

SCORES      Initl: 55 Initn: 55 Opt: 80 z-score: 125.0 E(): 3.2
Smith-Waterman score: 80; 23.5% identity in 81 aa overlap

pep2      30      40      50      60      70      80
      90      100      110      120      130      140
SLYVEKIDVGEAPRRTVSGLVQFVKPEELQDRLVVLVLCNLKPKKRGVES--QGMILCA
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
NCSNMIDEITTHLKQPLPLDFENNLNGEDODILMNNLRPNLEAFNRVKSQNNMS
10      20      30      40      50

pep2      90      100      110      120      130      140
      150      160
STEGINROYEPLDPAPGAPGEH-VFVKYEGKQPEELKPKKKYFEKQAQDFKISEECI
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
GIEAIRNLQPLCLPSATAPSRHPIIKAGDQGEFRKRTFYKLTLENQAQO
60      70      80      90      100      110

R78335      150      160
      10      20      30      40      50
AQMKTNFTKLGSISCKSLKGNIS

pep2
GENESQ:R56050
ID      R56050 standard; Peptide: 111 AA.
AC      R56050;
DE      23-JUN-1995 (first entry)
        Human Interleukin-3 mutain (pMon13350).

```

KW Human Interleukin-3; hIL-3; mutant; mutein.
OS Synthetic. . . .

SCORES Initl: 55 Initn: 55 Opt: 80 z-score: 125.0 E(): 3.2
Smith-Waterman score: 80; 23.5% identity in 81 aa overlap

pep2 SLVEKIDVGEAEPTVVSGLVQFVKELQDRLVVLCNLKPKKVGES--OGMLCA
R56050 NCSNMIDEITTHLKQPLPLDPNNLNGEDODILMENNLRPNLEAFNRAVKSQNAS
10 20 30 40 50 60 70 80

pep2 SIEGINROVEPLDPAGSAPGEH-VFVGKYGEGQDDEELKPKKVEKQLQADFKISEECI
R56050 GIEALRNLOPCLPSATAPSRHPITIKAGDWOEFRRKLTFTYLVLEQAQEOQ
60 70 80 90 100 110

pep2 AOWKQTNFMTKLGSISCKSLKGNIS

pep2
GENESEQP: P80381

ID P80381 standard; protein; 132 AA.
AC P80381;
DT 17-NOV-1990 (first entry)
DE Interleukin-3 variant;
KW Interleukin-3 variant; mutein;
KW colony stimulating factor activity. . . .

SCORES Initl: 44 Initn: 44 Opt: 81 z-score: 124.9 E(): 3.3
Smith-Waterman score: 81; 23.5% identity in 81 aa overlap

pep2 SLVEKIDVGEAEPTVVSGLVQFVKELQDRLVVLCNLKPKKVGES--OGMLCA
P80381 WVNCNMIDEITTHLKQPLPLDPNNLNGEDODILMENNLRPNLEAFNRAVKSQNAS
20 30 40 50 60 70 80

pep2 SIEGINROVEPLDPAGSAPGEH-VFVGKYGEGQDDEELKPKKVEKQLQADFKISEECI
P80381 AIESILKULPLCLPLATAPTRHPIHIDGDMNEPRKLTFTYLVLEQAQOQTLSLAI
80 90 100 110 120 130

pep2 AOWKQTNFMTKLGSISCKSLKGNIS

pep2
P80381 F

pep2
GENESEQP: P80383

ID P80383 standard; protein; 133 AA.
AC P80383;
DT 18-NOV-1990 (first entry)
DE Interleukin-3 variant;
KW Interleukin-3 variant; mutein;
KW colony stimulating factor activity. . . .

SCORES Initl: 44 Initn: 44 Opt: 81 z-score: 124.8 E(): 3.3
Smith-Waterman score: 81; 23.5% identity in 81 aa overlap

pep2 SLVEKIDVGEAEPTVVSGLVQFVKELQDRLVVLCNLKPKKVGES--OGMLCA
P80383 WVNCNMIDEITTHLKQPLPLDPNNLNGEDODILMENNLRPNLEAFNRAVKSQNAS
20 30 40 50 60 70 80

pep2 SIEGINROVEPLDPAGSAPGEH-VFVGKYGEGQDDEELKPKKVEKQLQADFKISEECI
P80383 AIESILKULPLCLPLATAPTRHPIHIDGDMNEPRKLTFTYLVLEQAQOQTLSLAI
80 90 100 110 120 130

pep2 AOWKQTNFMTKLGSISCKSLKGNIS

pep2
P80383 F

pep2
GENESEQP: R56095

ID R56095 standard; peptide; 113 AA.
AC R56095;
DT 26-JUN-1995 (first entry)
DE Human Interleukin-3 mutein (pMon13408).
KW Human Interleukin-3; hIL-3; mutant; mutein.
OS Synthetic. . . .

SCORES Initl: 54 Initn: 54 Opt: 80 z-score: 124.8 E(): 3.3
Smith-Waterman score: 80; 28.8% identity in 80 aa overlap

pep2 SLVEKIDVGEAEPTVVSGLVQFVKELQDRLVVLCNLKPKKVGESOGMLCAS
R56095 MANCSIMIDEITTHLKRPNPPLDPNNLNSSEDMIDILIRNRTPLLAFFVRAVKHLENAS
10 20 30 40 50 60 70 80

pep2 -IEGINROVEPLDPAGSAPGEH-VFVGKYGEGQDDEELKPKKVEKQLQADFKISEECI
R56095 GIEALRNLOPCLPSATAPSRHPITIKAGDWOEFRRKLTFTYLVLEQAQEOQ
70 80 90 100 110

pep2 AOWKQTNFMTKLGSISCKSLKGNIS

pep2
GENESEQP: R56105

ID R56105 standard; peptide; 113 AA.
AC R56105;
DT 26-JUN-1995 (first entry)
DE Human Interleukin-3 mutein (pMon13380).
KW Human Interleukin-3; hIL-3; mutant; mutein.
OS Synthetic. . . .

SCORES Initl: 54 Initn: 54 Opt: 79 z-score: 123.4 E(): 4
Smith-Waterman score: 79; 28.8% identity in 80 aa overlap

pep2 SLVEKIDVGEAEPTVVSGLVQFVKELQDRLVVLCNLKPKKVGESOGMLCAS
R56105 MANCSIMIDEITTHLKRPNPPLDPNNLNSSEDMIDILIRNRTPLLAFFVRAVKHLENAS
10 20 30 40 50 60 70 80

pep2 -IEGINROVEPLDPAGSAPGEH-VFVGKYGEGQDDEELKPKKVEKQLQADFKISEECI
R56105 GIEALRNLOPCLPSATAPSRHPITIKAGDWOEFRRKLTFTYLVLEQAQEOQ
70 80 90 100 110

pep2 AOWKQTNFMTKLGSISCKSLKGNIS

pep2
GENESEQP: R56098

pep2
GENESEOP:R56091
ID R56091 standard: peptide; 113 AA.
AC R56091;
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13440).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic. . . .

SCORES Initl: 54 Infn: 54 Opt: 77 z-score: 120.4 E(): 5.8
Smith-Waterman score: 77; 28.8% identity in 80 aa overlap

pep2
SLYVEKIDVGEAEPRIVVSGLVGFVPKEELDRLLVYLNLK-PQKMGVSGQMLTLCAS
30 40 50 60 70 80
MANCSIMIDEIHHLKRPNPDLDPNNLNSDDMDILMERNLRTPNLLAFVRAVHLENAS
10 20 30 40 50 60
pep2
-IEGINQVEPLDPAGSAPGEH-VFVKGYEKGQDEELPKKKVFEKLQADFRISECI
90 100 110 120 130 140
||:|:::| |::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
GIEAILNLQPCLPSTATAAPSRHPRIIKAGDMQEFREKLFYLVTLTQAOEQ
70 80 90 100 110
R56091

pep2
AQMKTNFMTRKLSISCKSLKGNIS

pep2
GENESEOP:R78352

ID R78352 standard: protein; 113 AA.
AC R78352;
DT 27-JUN-1996 (first entry)
DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
KW Interleukin-3; mutein; colony stimulating factor; CSF; chemotherapy;
radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody; . . .

SCORES Initl: 54 Infn: 54 Opt: 77 z-score: 120.4 E(): 5.8
Smith-Waterman score: 77; 28.8% identity in 80 aa overlap

pep2
SLYVEKIDVGEAEPRIVVSGLVGFVPKEELDRLLVYLNLK-PQKMGVSGQMLTLCAS
30 40 50 60 70 80
MANCSIMIDEIHHLKRPNPDLDPNNLNSDDMDILMERNLRTPNLLAFVRAVHLENAS
10 20 30 40 50 60
pep2
-IEGINQVEPLDPAGSAPGEH-VFVKGYEKGQDEELPKKKVFEKLQADFRISECI
90 100 110 120 130 140
||:|:::| |::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
GIEAILNLQPCLPSTATAAPSRHPRIIKAGDMQEFREKLFYLVTLTQAOEQ
70 80 90 100 110
R78352

pep2
AQMKTNFMTRKLSISCKSLKGNIS

pep2
GENESEOP:R56137

ID R56137 standard: peptide; 113 AA.
AC R56137;
DT 28-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13440).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic. . . .

SCORES Initl: 54 Infn: 54 Opt: 77 z-score: 120.4 E(): 5.8
Smith-Waterman score: 77; 28.8% identity in 80 aa overlap

30 40 50 60 70 80

pep2
SLYVEKIDVGEAEPRIVVSGLVGFVPKEELDRLLVYLNLK-PQKMGVSGQMLTLCAS
30 40 50 60 70 80
MANCSIMIDEIHHLKRPNPDLDPNNLNSDDMDILMERNLRTPNLLAFVRAVHLENAS
10 20 30 40 50 60
pep2
-IEGINQVEPLDPAGSAPGEH-VFVKGYEKGQDEELPKKKVFEKLQADFRISECI
90 100 110 120 130 140
||:|:::| |::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
GIEAILNLQPCLPSTATAAPSRHPRIIKAGDMQEFREKLFYLVTLTQAOEQ
70 80 90 100 110
R56137

pep2
AQMKTNFMTRKLSISCKSLKGNIS

pep2
GENESEOP:R78368

ID R78368 standard: protein; 113 AA.
AC R78368;
DT 27-JUN-1996 (first entry)
DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
KW Interleukin-3; mutein; colony stimulating factor; CSF; chemotherapy;
radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody; . . .

SCORES Initl: 54 Infn: 54 Opt: 77 z-score: 120.4 E(): 5.8
Smith-Waterman score: 77; 28.8% identity in 80 aa overlap

pep2
SLYVEKIDVGEAEPRIVVSGLVGFVPKEELDRLLVYLNLK-PQKMGVSGQMLTLCAS
30 40 50 60 70 80
MANCSIMIDEIHHLKRPNPDLDPNNLNSDDMDILMERNLRTPNLLAFVRAVHLENAS
10 20 30 40 50 60
pep2
-IEGINQVEPLDPAGSAPGEH-VFVKGYEKGQDEELPKKKVFEKLQADFRISECI
90 100 110 120 130 140
||:|:::| |::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
GIEAILNLQPCLPSTATAAPSRHPRIIKAGDMQEFREKLFYLVTLTQAOEQ
70 80 90 100 110
R78368

pep2
AQMKTNFMTRKLSISCKSLKGNIS

pep2
GENESEOP:R56120

ID R56120 standard: peptide; 113 AA.
AC R56120;
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13383).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic. . . .

SCORES Initl: 54 Infn: 54 Opt: 77 z-score: 120.4 E(): 5.8
Smith-Waterman score: 77; 31.3% identity in 83 aa overlap

pep2
DSLYVEKIDVGEAEPRIVVSGLVGFVPKEELDRLLVYLNLK-PQKMGVSGQMLTLCAS
30 40 50 60 70 80
MANCSIMIDEIHHLKRPNPDLDPNNLNSDDMDILMERNLRTPNLLAFVRAVHLENAS
10 20 30 40 50
pep2
CAS-IEGINQVEPLDPAGSAPGEH-VFVKGYEKGQDEELPKKKVFEKLQADFRISECI
90 100 110 120 130 139
||:|:::| |::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
NASGIEAILNLQPCLPSTATAAPSRHPRIIKAGDMQEFREKLFYLVTLTQAOEQ
60 70 80 90 100 110
R56120

pep2
ECIAQMKTNFMTRKLSISCKSLKGNIS

pep2

pep2
GENESEQP:R14839

ID R14839 standard; Protein; 464 AA.

ID	R14839	standard; Protein: 464 AA.
AC	R14839;	
DT	30-JAN-1992	(first entry)
DE	Protein deduced from human Glioblastoma cell cDNA	
KW	RAS; oncogene; cancer; cAMP; puc265.	
OS	Homo sapiens. . . .	

SCORES	Init1:	61	Initn:	61	Opt:	85	z-score:	119.6	E():	6.4
	Smith-Waterman score:	85;		28.2%	identity in	85	aa overlap			

pep2
R14839

40	50	60	70	80	90
VGEAEPRTVSGLVCPREELQDRILVYLCTNPKQKRGVSGMGLCASIEGINKROYE					
90	100	110	120	130	140
RRMAELSRDKCTVCGGLVDVYSLQENKECHVSTDTMLQTRIFQM--TVQKNLYLSSSS					

pep2 100 110 120 130 140 150
PLDPPAGS-APGEHVPVKSEKGPDEELKPKKVFEXLQADFRIISCEIAQWKQINPFMT
||||| - :::: - ||:: - ||| - | - :::: - ::|||
R14839 150 160 170 180 190
ELDPPISLIPEDQIDVV-LEKAMHKCIIPLKGAIVFEMLDIMAD--GSKQLKENL

pep2 160 KLGISICKSLKGNIS
R14839 200 QLVNRNQPEQLGYFAFTPDFVDERIKYKFMIMOKMSPERKVMILRLRVCKLIITYMENN
210 220 230 240 250

pep2	
GENESQ: R78367	
ID	R78367 standard; protein: 125 AA.
AC	R78367;
DT	27-JUN-1996 (first entry)
DE	Human IL-3 mutant for multi-lineage haematopoietic cell prod.
KW	Interleukin-3; mtleuh; colony stimulating factor; CSF; chemotherapy; .
RT	radiation; haematopoiesis; IL-3 antagonist; antileuk; antibody; . .

SCORES	Init1:	54	Initn:	54	Opt:	77	z-score:	119.5	E():	6.55
	Smith-Waterman score:	77;		28.8%	identity in	80	aa overlap			

pep2
R78367

30 40 50 60 70 80
SLYERIDVGEAPRTVSGLVQFVKEELDRLLVYLNLK-PQKRGVESQGLMCAS
DKNCIMDEILTHLKRPNPLDNNLNSEMDILMERNLRTNLLAFRAVKHLENS

pdp2
R78367

	90	100	110	120	130	140
-	TGIRQRYEPLDP	PAGSGA	PGEH-VFV	KGYEKQDDE	LKKKVFVE	LOADRTISECT
	: :	:	:	:	:	:
G	EALIRNCPCL	SPSTAPSR	RHPITII	IAGWQDER	REKLTFY	LTLOAOEQO
	:	:	:	:	:	:
80	90	100	110	120		

pep2 150 160
GENESE00:R56082 A Q M K Q T N F M T K L G S I S C K S L K G C N I S

ID	Accession	Protein	Gene	Species	Length (aa)	Weight (kDa)	pI	Inst.	Ref.
R56082	standard; peptide; 126 AA.								
AC	R56082;								
DT	30-JUN-1995 (first entry)								
DE	Human interleukin-3 mutein (PM013330).								
KW	Human interleukin-3; hIL-3; mutant; mutein								
OS	Synthetic.								

SCORES	Initl:	54	Inltn:	54	Opt:	77	z-score:	119.4	E():	6.6
Smith-Waterman score:	77;		28.88	identity	in	80	aa overlap			

	10	20	30	40	50	60	70	80	90	100
pep2	SLYEKIDNGAEAPRVVSGVLQVAPKEELODRLLVYLCLNIK	-	1	:	:	:	:	:	:	:
R56082	KANCINIDELIHHLKRPENPLLDNNININSEMDIMLRNLRTPENLLAFVAVKHLNENSA	-	:	:	:	:	:	:	:	:
	20	30	40	50	60	70				
pep2	-IGSINQVPEPLDPRGASAPGEH-VFVAGYEGQPDDELAKKKVFFEQKADRIISBECL	-	:	:	:	:	:	:	:	:
R56082	GIEILNINLOPCLPSTAPASHPIILIKAGDWQERREKLEIFYLVLTLEQAOEQQ	-	:	:	:	:	:	:	:	:
	80	90	100	110	120	130	140			

pep2 AQQKQTNFMTKGLSISCKSLKGNIS
pep2
GENESEQOP:R72583

ID R72583 standard; peptide; 15 AA.
AC R72583;
DT 29-SEP-1995 (first entry)
DE EMAP1 active peptide.
KW EMAP1; endothelial monocyte activating polypeptide II; chemotaxis
inflammation; tissue factor; tumor; cancer; therapy; metha; . . .

SCORES	Initl:	25	Initn:	25	Opt:	64	z-score:	119.2	E():	6.8
	Smith-Waterman score:	64;		57.1%	identity in	14	aa overlap			

pep2
R72583

10 20 30 40 50 60
EEVIPSRLDRGKITTVEKHPDASLVEKNIDVGEALPRIVSGLVQFVFKPELDRLVY
||||:|:|:|:
ASRLDRIGRIYAK
10

```

pep2
GENESECP:R56093

ID   R56093 standard; peptide; 113 AA.
DC   R56093:
DT   26-JUN-1995 (first entry)
DE   Human interleukin-3 mutein (Pmon13405).
KW   Human interleukin-3, bIL-3, mutant; mutein
     Synthetic. . .

```

```

SCORES      Initl:  54  Initn:  54  Opt:  76  z-score: 118.9  E():  7
Smith-Waterman score: 76;  28.8% identity in 80 aa overlap

```

pep2 30 40 50 60 70 80
 SLVKEKIDVGEAPRVVSGLVQFPEKEDLDRLVYVNLK-PQKRGVESQGMGLAS
 R56093 10 20 30 40 50 60
 MANCSIAIDELIHLKRPNPPLDPPNNNSDVIDILMERIKRTPMLAFLRAVAKHLNNS

```

90      100      110      120      130      140
pep2    -IGIRNOVELDPPGASNGEH-VFVKGKYGQDDELPRKKVKFEKLOADRKISECT
          |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
R56093  GIEALIRNLOPCLPSATAPSRNPIITAGWQEFKRLFFYVLTLEQAEQQ
          70      80      90      100      110

```

pep2	150	160
GENESEQP:R56068	ADWKQTFMTKLGSISCKSLKGNIS	

ID	R56068	standard; peptide; 113 AA.
AC	R56068;	
DT	22-JUN-1995	(first entry)

DE Human Interleukin-3 mutein (pMon13289).
 KW Human Interleukin-3; hIL-3; mutant; mutein.
 OS Synthetic. . . .

SCORES Initl: 54 Initn: 54 Opt: 76 z-score: 118.9 E(): 7
 Smith-Waterman score: 76; 28.8% identity in 80 aa overlap

pep2 SLVEKIDVGEAEPRIVVSGLVQFVPEKELODRLVVVLGNLK-PQKMGVESQGMILCAS
 30 40 50 60 70 80
 R56068 MANCSIMDELIHHLKVPAPLDSNNINSEDMQIMERNLRLPLLFVFAVAKLENAS
 10 20 30 40 50 60

pep2 -IEGINRQVEPLDPAGSAPGHEH-VFVKGYEKGQDPEELPKKKVFEKLQADFKISEECI
 90 100 110 120 130 140
 R56068 GIEALIRNLQPCLPASATAPSRHRPIITIKAGDMQEFREKLTFTYVLTLEAQEQ
 70 80 90 100 110

pep2 AQMKTNFMTRLKGISCSKSLKGNIS

pep2
 GENESQ:R56092

ID R56092 standard; peptide; 113 AA.

AC R56092;

DT 26-JUN-1995 (first entry)

DE Human Interleukin-3 mutein (pMon13407).

KW Human Interleukin-3; hIL-3; mutant; mutein.

OS Synthetic. . . .

SCORES Initl: 54 Initn: 54 Opt: 76 z-score: 118.9 E(): 7
 Smith-Waterman score: 76; 28.8% identity in 80 aa overlap

pep2 SLVEKIDVGEAEPRIVVSGLVQFVPEKELODRLVVVLGNLK-PQKMGVESQGMILCAS
 30 40 50 60 70 80
 R56092 MANCSIMDELIHHLKRPNPDLDPNNINSEDMQIMERNLRLPLLFVFAVAKLENAS
 10 20 30 40 50 60

pep2 -IEGINRQVEPLDPAGSAPGHEH-VFVKGYEKGQDPEELPKKKVFEKLQADFKISEECI
 90 100 110 120 130 140
 R56092 GIEALIRNLQPCLPASATAPSRHRPIITIKAGDMQEFREKLTFTYVLTLEAQEQ
 70 80 90 100 110

pep2 AQMKTNFMTRLKGISCSKSLKGNIS

pep2
 GENESQ:R56101

ID R56101 standard; peptide; 113 AA.

AC R56101;

DT 26-JUN-1995 (first entry)

DE Human Interleukin-3 mutein (pMon13425).

KW Human Interleukin-3; hIL-3; mutant; mutein.

OS Synthetic. . . .

SCORES Initl: 54 Initn: 54 Opt: 76 z-score: 118.9 E(): 7
 Smith-Waterman score: 76; 28.8% identity in 80 aa overlap

pep2 SLVEKIDVGEAEPRIVVSGLVQFVPEKELODRLVVVLGNLK-PQKMGVESQGMILCAS
 30 40 50 60 70 80
 R56101 MANCSIMDELIHHLKRPNPDLDPNNINSEDMQIMERNLRLPLLFVFAVAKLENAS
 10 20 30 40 50 60

pep2 -IEGINRQVEPLDPAGSAPGHEH-VFVKGYEKGQDPEELPKKKVFEKLQADFKISEECI
 90 100 110 120 130 140
 R56101 GIEALIRNLQPCLPASATAPSRHRPIITIKAGDMQEFREKLTFTYVLTLEAQEQ
 70 80 90 100 110

pep2 AQMKTNFMTRLKGISCSKSLKGNIS

! CPU time used:
 ! Database scan: 0:01:37.3
 ! Post-scan processing: 0:00:02.9
 ! Total CPU time: 0:01:40.2
 ! Output file: pep2.fasta
 ! IAA_SEQUENCE 1.0

ID W06596 standard; Protein; 168 AA.

AC W06596;

DT 04-JUL-1997 (first entry)

DE Human endothelial-monocyte activating polypeptide III.

KW Endothelial-monocyte activating polypeptide; EMAP III; human;

KW neoplasia.

OS Synthetic.

PN W09640719-A1.

PD 19-DEC-1996.

PE 07-JUN-1995; U07328.

PR 07-JUN-1995; WO-U07328.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Coleman TA, Olsen HS, Rosen CA;

DR WPI: 97-052219/05.

DR N-PSDS; T49312.

PT New isolated endothelial-monocyte activating polypeptide III - used

PT to develop prods. for the diagnosis and treatment of conditions

PT involving EMAP III, e.g. neoplasia

PS Claim 4: Page 38; 50pp; English.

CC The present sequence represents the 168 amino acid sequence for the novel

CC polypeptide endothelial-monocyte activating polypeptide III (EMAP III).

CC The product can be used for the diagnosis and treatment of conditions

CC involving abnormal expression of EMAP III. In particularly EMAP III can

CC be used to regress neoplasia such as in tumours and cancers.

CC EMAP III shows a high degree of homology to EMAP II with 60%

CC identity and 75% similarity over a 150 amino acid stretch.

SQ Sequence 168 AA;

SQ 7 A; 6 R; 4 N; 8 D; 0 B; 4 C; 9 Q; 18 E; 0 Z; 13 G; 2 H;

SQ 11 I; 14 L; 18 K; 3 M; 5 F; 11 P; 11 S; 4 T; 1 W; 2 Y; 17 V;

W06596 Length: 168 April 28, 1998 12:48 Type: P Check: 3813 ..

1 EEVIPSRLDI RVGKITVER HPDADSLYE KIDVGEAEPR TVVSGLVQFV

51 PREELQRLV VVLGNLKPQK MGVESQGM LCAISIEGINR QVEPLDPAG

101 SAGEHVFVK GYEKGQDPEE LKPKKVFKE LQADFRISEE CIQMKTNFM

151 MTKGISCSK SLKGNIS

! IAA_SEQUENCE 1.0

ID W14561 standard; protein; 166 AA.

AC W14561;

DT 01-DEC-1997 (first entry)

DE Endothelial monocyte activating polypeptide II.

KW Endothelial monocyte activating polypeptide; blood vessel;

KW EMAP-II; inhibitor; endothelial cell formation; blood vessel;

KW retinopathy; tumour; subcutaneous; intraperitoneal; intravenous;

KW intracranial; intratumoural; carcinoma; glioblastoma;

KW positive pressure microinfusion; aortic cell; angiogenesis;

KW eye disease; diabetes; sickle cell anaemia; prematurity;

KW age-related macular degeneration.

OS Mus musculus.

FT key location/Qualifiers

FT Misc_difference 1

FT /label- Ser, Met, Gly

FT Misc_difference 15

FT /label- Cys, Arg

PN W09710841-A1.
 PD 27-MAR-1997.
 PR 18-SEP-1996; U15007.
 PR 18-SEP-1995; US-003898.
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 PI Schwarz M, Stern D;
 DR WPI: 97-202617/18.
 PT Treatment of tumours with endothelial monocyte activating peptide II
 PT - also used to inhibit growth of endothelial cells and formation of
 PT blood vessels, e.g. in retinopathy
 PS Claim 13; Page 73; 104pp; English.
 CC A method has been developed for the treatment of tumours. The method
 CC involves subcutaneous, intraperitoneal, intravenous, intracranial or
 CC intratumoural administration of an endothelial monocyte activating
 CC polypeptide II (EMAP-II), or a EMAP-II-derived polypeptide. The present
 CC sequence represents murine EMAP-II which can be used to produce the
 CC EMAP-II-derived polypeptide as long as it contains at least ninety
 CC percent homology to the present sequence. The method is especially used
 CC to treat carcinoma (administered intraperitoneally) or glioblastoma
 CC (administered intratumourally by positive pressure microinfusion).
 CC Inhibition of endothelial cell growth is particularly applied to aortic
 CC cells and inhibition of angiogenesis to treatment of eye diseases
 CC associated with excessive blood vessel formation, especially retinopathy
 CC (diabetic, related to sickle cell anaemia, retinopathy of prematurity or
 CC age-related macular degeneration). Intraperitoneal administration allows
 CC tumours too small (not over 2 mm in diameter) for intratumoural
 CC injection to be treated before they grow larger.
 SO Sequence 166 AA;
 SO 15 A; 7 R; 7 N; 9 D; 0 B; 4 C; 6 Q; 11 E; 0 Z; 11 G; 3 H;
 SO 8 I; 12 L; 14 K; 6 M; 3 F; 14 P; 8 S; 6 T; 1 W; 2 Y; 17 V;
 SO 2 Others;

W14561 Length: 166 April 28, 1998 12:48 Type: P Check: 5793 ..

1 NKPIASRLD LRIGNITYAK KHPDADSLY EEDVNGEAP RTVSGLYNH
 51 VPLEQONRM VLLCNLKP KMGVLSQAM VMCASSPEV EILAPNGSV
 101 PGDRITFDNF PGEPDKELNP KKKIMEQIOP DLHTNAEVA TYKGAPFEVK
 151 GKGVRAQTM ANSGIK
 !!AA_SEQUENCE 1.0
 ID R72577 standard; Protein: 310 AA.
 AC R72577;
 DT 29-SEP-1995 (first entry)
 DE Mouse EMAP-II.
 KW EMAP-II; endothelial monocyte activating polypeptide II; Chemotaxis;
 KW Inflammation; tissue factor; tumor; cancer; therapy; metha;
 KW sarcoma.
 OS Mus musculus.
 PN W09509180-A.
 PD 06-APR-1995.
 PR 29-SEP-1994; U11085.
 PR 29-SEP-1993; US-129456.
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 PI Clauss M, Kao J, Kayton M, Libutti SK, Stern DM;
 DR WPI: 95-147389/19.
 DR N-PSDB; Q86718.
 PT New endothelial monocyte activating polypeptide II - induces
 PT chemotaxis, inflammation and tissue factor, useful for treating
 PT tumours, also related antibodies, DNA and active fragments
 PS Disclosure: Fig.4; 180pp; English.
 CC A mouse metha sarcoma cDNA library was screened with a probe
 CC based on the N-terminal sequence of mouse EMAP-II. Overlapping
 CC clones were combined to obtain a contiguous full-length sequence
 CC (given in Q86718) encoding a 33 kDa protein (R72577).
 CC Recombinant EMAP-II was expressed in E. coli.
 SO Sequence 310 AA;
 SO 32 A; 12 R; 11 N; 12 D; 0 B; 6 C; 18 Q; 32 E; 0 Z; 15 G; 4 H;
 SO 15 I; 25 L; 36 K; 8 M; 3 F; 17 P; 17 S; 16 T; 1 W; 3 Y; 27 V;
 R72577 Length: 310 April 28, 1998 12:48 Type: P Check: 1208 ..

1 MATNDVILKR LEQGAEDQ IIEYKQYA LKEKAIQA TMREKKLRV
 51 ENAKLKEIE ELKEELIAE IHNGVQAVY RLTSLQINC TASSVQSP
 101 SVATTASPAT KEQIKAGEER KYEKTEKKG EKKEKQSSAA ASTDSKPDA
 151 SRLDRICGI VTAKKHPDAD SLVEEDVG EAAPRTVSG LVNHVPLEOM
 201 QNRMYVILCN LKPAKMGVY SQAMWCASS PEKVELIAP NGSPGDRIT
 251 FDAFGEEDK ELNPKKIWE QIOPDLHTNA ECVATYKGP FEVKGVCYR
 301 AQTMANSGBK
 !!AA_SEQUENCE 1.0
 ID W20051 standard; Protein: 657 AA.
 AC W20051;
 DT 10-SEP-1997 (first entry)
 DE Methionyl-tRNA synthetase from Staph. aureus.
 KW tRNA synthetase; Bacillus subtilis; mets; immunological response;
 KW antibody; bacterial infection; adherence; damaged tissue;
 KW wound healing; skin; protection; vaccine.
 OS Staphylococcus aureus.
 PD EP-785268-A1.
 PR 23-JUL-1997.
 PR 17-JAN-1997; 300317.
 PR 19-JAN-1996; GB-001095.
 PR 27-JUL-1996; GB-015845.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PI Hodgson JE, Lawlor EJ;
 DR WPI: 97-365943/34.
 DR N-PSDB; T71309.
 PT DNA encoding methionyl-tRNA synthetase from Staphylococcus aureus
 PT WCUH 29 - useful for protection against bacterial infections
 PS Claim 1; Pages 19-20; 31pp; English.
 CC The present sequence is a novel methionyl tRNA synthetase
 CC protein from staphylococcus aureus strain WCUH29 (NCIMB 40771),
 CC which is related by amino acid sequence homology to Bacillus
 CC subtilis tRNA synthetase encoded by the mets gene. The enzyme
 CC catalyses the aminoacylation of tRNA-Met, by a two step mechanism.
 CC The first step involves formation of a stable enzyme:methionyl
 CC adenylylate complex resulting from the specific binding and reaction
 CC of ATP and L-methionine. Subsequently, the 3' terminal adenosine of
 CC enzyme-bound tRNA-Met reacts with the aminoacyl adenylylate, leading
 CC to the esterification of the tRNA and release of AMP. Vectors
 CC comprising the DNA (or polynucleotides having at least 70 %
 CC identity to it) can be used for the recombinant production of the
 CC enzyme. The enzyme or its related DNA (through gene therapy) is
 CC used to induce an immunological response in a mammal to generate
 CC antibodies to protect against disease. The antibodies protect
 CC against invasion of bacteria, e.g. by blocking adherence of
 CC bacteria to damaged tissue, including wounds in skin or connective
 CC tissue caused by mechanical, chemical or thermal damage or by
 CC implantation of in-dwelling devices, or wounds in the mucous
 CC membranes.
 SO Sequence 657 AA;
 SO 40 A; 20 R; 27 N; 43 D; 0 B; 0 C; 24 Q; 55 E; 0 Z; 36 G; 15 H;
 SO 51 I; 57 L; 51 K; 20 M; 26 F; 35 P; 41 S; 31 T; 9 W; 32 Y; 44 V;
 W20051 Length: 657 April 28, 1998 12:48 Type: P Check: 1909 ..
 1 MAKEFFIYTT PIYPSGNLH IGHAYSTVAG DVIARKKMQ GYDVRILTGT
 51 DEHGKRIQEK AQKAGKTEIE YLDEMIAGIK QUMAKLEISN DDFRTTEER
 101 KHVVEQVFE RLKQGDYIL GEYEGWVSV DDTYTESQL VDPQYENKXI
 151 ICGKSPDSGH EVELYKEESY FFINISKYTDL LLEFDQNDP FIQPPSKANE
 201 MINNFIKGL ADLAVSRTSF NMGVHPSNP KHVVYVWIDA LVNYSALGY
 251 LSDDSLFLNK YMPADIHMA KEIVRFHSII WILMALDUL PLPKVFPAHG

301 WILMKDGKMS KSKGNVDPN IIDRYGLDA TRYLMRELPG FSGDGVETPE
351 AFVERTNEPL ANDLGNLVNR TISMVKNYFD GELPAYOGPL HELDEMEAM
401 ALEEVKASTE SMESLOSVA LSTVWKFISR TNKYIDETTP WYLAKDSOK
451 DMLGNVAHL VENIRYRAVL LRPELTNAPK EIFEOLNINN PQEWEFSLE
501 QYGVLTESIM VTGQPKPIFP RLDSEALIAV IKESMQPPAT EEEKELPSK
551 PQDIKQEDK VEIKAAIID AEVKKSDKL LKIQVDLDE QROIVSGIAK
601 FYTPDILIGK KVAVTNLKP AKLMQKSEK MILSAEKDGV LTVLSLPSAI
651 PNGAVIK

11AA_SEQUENCE 1.0
ID W20570 standard; Protein; 436 AA.
AC W20570;
DT 17-JUL-1997 (first entry)
DE H. pylori cytoplasmic protein 785437.aa.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS Helicobacter pylori.
PN MO9640893-A1.
PD 19-DEC-1996.
PE 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR) ASTRA AB.
PI Berglindh OJ, Smith D, Mellgaard BJ;
DR WPI: 97-052306/05.
DR N-PSDB: T67713.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 61, Page 724-725; 1481pp; English.
CC This sequence is a H. pylori cytoplasmic protein involved in
CC mRNA translation and ribosome biogenesis.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 436 AA;
SQ 35 A; 15 R; 28 N; 17 D; 0 B; 2 C; 12 Q; 42 E; 0 Z; 19 G; 8 H;
SQ 30 I; 61 L; 48 K; 14 M; 18 F; 15 P; 23 S; 8 T; 5 W; 15 Y; 21 V;
SQ

W20570 Length: 436 April 28, 1998 12:48 Type: P Check: 8042 ..

1 MNDKRVVYV WLDALNTAS ALGYLGLDN KMAHERARH IVGKILIRFH
51 AIWPAFLMS LNLPLFKOLC VHGWMTEGV KMSKSLGNVL DAQKLAIEYG
101 IEEARVFLR EVPEFGDGF SKKALVERIN ANLNNDLGNL LNRLLGAKK
151 YFVNSLSTK ITAIYPRELE KAHQIIDNAN SFVPRMQLKH ALEELFNITD
201 FLNKLIAKEE PAVLHKNNES EKLEALLSLI ANTLLOSSF LVAFMRPSAM
251 KLASAFNVEI TPNNYERFFK AKKLQDMVQ DTEPLFSKIE KIEKIEIKK
301 IEKIEKEEBA LAEKAEKKEK EKAPTOENY ISIEDFKKVE IKVGLIKEAQ

351 RIEKSNKLLR LKVDLGENRL RQIISGIALD YEPESLVGOM VCYVANKKPA
401 KLMGMESEGM IIAVRNDNL ALISPTREKI AGSLIS

11AA_SEQUENCE 1.0
ID W21896 standard; Protein; 800 AA.
AC W21896;
DT 10-SEP-1997 (first entry)
DE Phenylalanyl tRNA synthetase beta subunit.
KW tRNA synthetase; Bacillus subtilis; immunological response; antibody;
KW bacterial infection; adherence; damaged tissue; wound healing; skin;
KW protection; vaccine.
OS Staphylococcus aureus.
PN EP-785262-A1.
PD 23-JUL-1997.
PE 17-JAN-1997; 300311.
PR 19-JAN-1996; GB-001096.
PR 27-JUL-1996; GB-015845.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Hodgson JE, Lawlor EJ;
DR WPI: 97-365937/34.
DR N-PSDB: T73694.
PT DNA encoding phenylalanyl-tRNA synthetase alpha subunit from
PT Staphylococcus aureus WCUH 29 - useful for protection against
PT bacterial infections
PS Disclosure: Page 22-24; 41pp; English.
CC The present sequence is a novel phenylalanyl tRNA synthetase beta subunit
CC from Staphylococcus aureus strain WCUH29 (NCIMB 40711), which is related
CC by amino acid sequence homology to Bacillus subtilis phenylalanyl tRNA
CC synthetase beta subunit. The enzyme catalyses the aminoacylation of
CC tRNA-Phe, by a two step mechanism. The first step involves formation
CC of a stable enzyme: phenylalanyl adenylate complex resulting from the
CC specific binding and reaction of ATP and L-phenylalanine. Subsequently,
CC the 3' terminal adenosine of enzyme-bound tRNA-Phe reacts with the
CC aminoacyl adenylate, leading to the esterification of the tRNA and
CC release of AMP. Vectors comprising the DNA (or polynucleotides having at
CC least 70% identity to it) can be used for the recombinant production of
CC the enzyme. The enzyme or its related DNA (through gene therapy) is used
CC to induce an immunological response in a mammal to generate antibodies to
CC protect against disease. The antibodies protect against invasion of
CC bacteria, e.g. by blocking adherence of bacteria to damaged tissue,
CC including wounds in skin or connective tissue caused by mechanical,
CC chemical or thermal damage or by implantation of in-dwelling devices,
CC or wounds in the mucous membranes.
SQ Sequence 800 AA;
SQ 54 A; 38 R; 35 N; 64 D; 0 B; 4 C; 33 Q; 63 E; 0 Z; 52 G; 11 H;
SQ 67 I; 72 L; 42 K; 15 M; 25 F; 28 P; 49 S; 48 T; 3 W; 28 Y; 69 V;
SQ

W21896 Length: 800 April 28, 1998 12:48 Type: P Check: 2312 ..

1 MLINENLKE YVITIDSVSD LAERTIRGI EVDLDIYTK DINLVVGFV
51 KSKKEHPDAD KLVNCOVDIG EDEPVQIVCG APNVDAQVY IVAKVGRLP
101 GGIKIKRAKL RGERSEGMIC SLQEGISSN YIPKSFESGI YVESQVPG
151 TDAQALYLD DQWMEFDLTP NRADALSMIG TAYEVALYN TKMTKPEITS
201 NELELSANDE LVTIENEDK VPIYSARVH DVITIEPPIW MQARLIKAGI
251 RPINNVDIS NYVILEYGOP LHMEDQDAIG SQQIVVRQAN EGEKMTLLD
301 TERELLTSDI VITNGOTPIA LAGVAGGQFS EVKEQTSNIV IEBAINDPVS
351 IRTSRLNL RSESSRFEK GIATEFVDEA VDRACYLLQT YANGKVLKDR
401 VSSGELGAFI TPIDITADKI NRTIGFDSQ NDIYTFINOL GFDTETINDV
451 ITIVPSRRK DITIKEDLIE EVARIYGYD IPSTLPYFDK VISGQULDRQ
501 YTRMVKREV LAGAGLDOAIT YSLVSKEDAT AFSMQOROTI DILMPNSEAH

```

551 ASLSQSLPH LIEVASTINVA RKNKVLFIE IGWFFANGE GELPDQVEYL
601 SGILTGDIYV NQWQKKETV DEFLAKGYVD RVSEKLNLEF SYRRADIDGL
651 HPGRTAEILL ENKVYGEIGE LHPITLADND LKRTYVFELN FDLAMAVSG
701 YINQPIPRF PGMSRDIALE VDQNIIPADL LSTIHAGGN ILKDTLVEDV
751 YQGEHLKRGK KSIAIRLNTL DTEETLTDER VSKVQAEIEA ALIDQAVIR

!!A_SEQUENCE 1.0
ID W20999 standard: protein; 662 AA.
W20999;
DT 21-JUL-1997 (first entry)
DE H. pylori cytoplasmic protein, hp414522orf11.
KM Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KM binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS Helicobacter pylori.
PN W09640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996: U09122.
PR 07-JUN-1995: US-487032.
PR 01-APR-1996: US-630405.
PA (ASTR ) ASTRA AB.
PI Bergljindh Or, Smith D, Mellgaard BL;
PI WPI: 97-052306/05.
DR N-PSDB: T68252.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 61, Page 1389-1390: 1481pp: English.
CC The present sequence is a H. pylori cytoplasmic protein involved in
CC mRNA translation and ribosome biogenesis.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 662 AA;
SQ 46 A; 21 R; 37 N; 30 D; 0 B; 8 C; 21 Q; 60 E; 0 Z; 30 G; 12 H;
SQ 48 I; 80 L; 66 K; 16 M; 31 F; 24 P; 39 S; 26 T; 7 W; 31 Y; 29 V;
SQ

W20999 Length: 662 April 28, 1998 12:48 Type: P Check: 5242 ..
1 GNTIIMQKS LITPIIYV DIPHGHAYT TLIDTLKKY YTLQGEVFF
51 LTGTDEHGOK IEQSARLBNQ SPKAVADSTS ALFNQWDF NLDYDGIRI
101 TDSHQVCQV NAFIMPEKG DIYKGTISY YCVSCESCVA VSKVNDTDSK
151 VLPDCLRET TLLEESYFF KLSAYEKPL EFYAKNPRAI LPIYRKNQEV
201 SFIEGGLDL SITRTSEWG IPLPKMNDP KHVYVWMDA LNVASALGY
251 LINGDNMAH FERARIVGK DILRFHAIVY PAFMLSLNLP LFKQLCVHGW
301 WTIEGVMSK SLGNVDAQK LAMEYGIEL RYFLIREVPF GODGDEFSKA
351 LVERINANLN NDGNLNLRL LGMAKKYFN SLKSTKIYAY YPKLELAHQ
401 ILDNANFVP KMLHKALEE LENITDFLAK LIAKEEPVVL HKNNESEKLE
451 ALLSIANTL LOSSFLIYAF MPKSAKLAS AFREITPPNN YERFFKAKL

```

```

501 QDWVLDQTEP LFSKIEKIEK IEKIEKIEK EKGEEALAEK AERKEKEKAP
551 PTQENYISIE DFKKVEIKV LKEAQRIEK SNKLLRLKVD LGENRLROI
601 SGILDYEPE SLVGQWCVV ANLKPAKLMG EMSEGMIIAV RDNDAIALIS
651 PTRREKISL IS

!!A_SEQUENCE 1.0
ID R13875 standard: protein; 248 AA.
R13875;
DT 11-OCT-1991 (first entry)
DE Murine Cytotoxic Cell Protease-2.
KM mouse; CCP2 inhibitor; cytotoxic T-lymphocytes; ss.
OS Mus musculus.
PN W09110685-A.
PD 25-JUL-1991.
PF 17-JAN-1991: U00340.
PR 19-JAN-1990: US-467880.
PA (SERA-) SERAGEN INC.
PI Bleackley RC, Lobe CG, Paetkau VH, James MN, Murphy M;
PI WPI: 91-23789/32.
DR N-PSDB: Q12863.
PT DNA vectors, and inhibitors of cytotoxic cell protease - for
PT treatment of auto-immune diseases e.g. pernicious anaemia,
PT rheumatoid arthritis, allo-graft rejection etc.
PS Claim 6, Fig 10: 62pp: English.
CC The CCP2 coding sequence was isolated from the cytotoxic T-cell
CC line MTL 2.8.2 generated from CBA/J mice. The amino acid sequence
CC of CCP2 was predicted from the cDNA sequence. The structure of the
CC protein was used to design peptides which competitively inhibit the
CC protease. See Q12862-6 and R13254-R13262.
SQ Sequence 248 AA;
SQ 19 A; 13 R; 9 N; 11 D; 0 B; 6 C; 8 Q; 13 E; 0 Z; 20 G; 7 H;
SQ 13 I; 20 L; 21 K; 6 M; 9 F; 16 P; 16 S; 12 T; 2 W; 7 Y; 20 V;
SQ

R13875 Length: 248 April 28, 1998 12:48 Type: P Check: 6603 ..
1 MPVLILITL LPLRGAEE IIGNEISPH SRPYMYEYF LKYGKKMFC
51 GGLVRDKFV LTAHCKGSS MTVLGAHNI KAKEEQOII PVAKAIPHPD
101 YNPDRSNDI MLKLVRNAK RTRAVRPLN PRRAHVKPG DECVYAGWK
151 VTPGCEPKT LHEVKTLYOK DOVCESOFOS SYNANEICV GDSKXIGASF
201 EEDSGPLVC KRAAGIVSY GQTDGAPQV FTRVLSEVSW IKRTMRHS

!!A_SEQUENCE 1.0
ID R42379 standard: protein; 176 AA.
R42379;
DT 19-APR-1994 (first entry)
DE Haemin-binding protein produced as fusion protein in pCCH5.
KM Haemophilus somnus; immunogenic; haemolysin; lppb; lppc;
KM thromboembolic meningencephalitis; septicaemia; arthritis;
KM pneumonia; Ikta gene; Pasteurella haemolytica.
OS Haemophilus somnus.
PN W09321323-A.
PD 28-OCT-1993.
PF 05-APR-1993: CA0135.
PR 09-APR-1992: US-865050.
PR 04-JUN-1992: US-893424.
PR 04-JUN-1992: US-893426.
PR 29-MAR-1993: US-038287.
PR 29-MAR-1993: US-038288.
PR 29-MAR-1993: US-038719.
PA (UYSA-) UNIV SASKATCHEWAN.
PI Harland RJ, Pfeiffer CG, Pontarollo RA, Potter AA;
PI Rioux C, Theisen M;
DR WPI: 93-351733/44.
DR N-PSDB: Q51081.
PT Haemophilus somnus immunogenic proteins used in vaccines -
PT selected from haemin-binding protein, haemolysin, lppb and lppc,

```

PT and corresp. DNA
 PS Claim 1; Fig 5; 119pp; English.
 CC The hmb gene encoding the haemin-binding protein was expressed in
 CC E. coli as a fusion to the Pasteurella haemolytica leukotoxin gene
 CC lktA coded for by plasmid pHA352. The hmb gene fragment was taken
 CC from pRAP501 and starts at the codon for the third amino acid residue
 CC of ORF1. The haemin binding protein can be used in vaccines for
 CC preventing or treating H. somnus infections, which cause thromboembolic
 CC meningo-encephalitis, septicæmia, arthritis and pneumonia in
 CC vertebrates.
 CC See also R42370-86.
 SQ Sequence 176 AA;
 SQ 16 A; 10 R; 10 N; 8 D; 0 B; 7 C; 11 Q; 9 E; 0 Z; 18 G; 4 H;
 SQ 10 I; 14 L; 14 K; 1 M; 7 F; 5 P; 6 S; 7 T; 3 W; 6 Y; 10 V;

R42379 Length: 176 April 28, 1998 12:48 Type: P Check: 5317 ..

1 EFTKWTGAI CSVAIIALV KAHOELEIS QOGLDLIGNV EECRRDPYHC
 51 PADVLTVGIG STEANGKNID PKRYSDEI AQWYDRL AEQVARYGN
 101 GKNLPQAGD AFSITFNVG CGKQKSTLF KOANOGFTPO LCHOFRWY
 151 AGGKTLGLV ARRAKALC LGEYHD

11AA_SEQUENCE 1.0
 ID R42374 standard; Protein: 178 AA.

AC R42374;
 DT 19-APR-1994 (first entry)
 DE Haemin-binding protein from ORF5 of plasmid pRAP501.
 DE Haemophilus somnus; immunogenic; haemolysin; LppB; LppC;
 KW thromboembolic meningoencephalitis; septicæmia; arthritis;
 KW pneumonia; haemin-binding protein.
 OS Haemophilus somnus.
 PN WO9321323-A.
 PD 28-OCT-1993.
 PF 05-APR-1993; CA0135.
 PR 09-APR-1992; US-865050.
 PR 04-JUN-1992; US-893424.
 PR 04-JUN-1992; US-893426.
 PR 29-MAR-1993; US-038287.
 PR 29-MAR-1993; US-038288.
 PR 29-MAR-1993; US-038719.
 PA (UYGA-) UNIV SASKATCHEWAN.
 PI Hatland RJ, Pfeiffer CG, Pontarollo RA, Potter AA;
 PI Rioux C, Theisen M;
 PI MPI: 93-351733/44.
 DR N-PSDB: Q51080.
 DR Haemophilus somnus immunogenic proteins used in vaccines -
 PT selected from haemin-binding protein, haemolysin, LppB and LppC,
 PT and corresp. DNA

PS Claim 1; Fig 2; 119pp; English.
 CC A genomic cosmid library of Haemophilus somnus HS25 DNA was screened
 CC for clones capable of binding bovine haemin and having haemolytic
 CC activity. Positive clones were subcloned various times, resulting
 CC in plasmid pRAP501, which binds haemin but is not haemolytic. The
 CC clone was sequenced and was found to contain several open reading
 CC frames, potentially encoding 8 proteins. The haemin binding protein
 CC (encoded by the hmb gene) was encoded by ORF1. The protein can be
 CC used in vaccines for preventing or treating H. somnus infections,
 CC which cause thromboembolic meningo-encephalitis, septicæmia,
 CC arthritis and pneumonia in vertebrates.
 CC See also R42370-86.
 SQ Sequence 178 AA;
 SQ 16 A; 10 R; 11 N; 8 D; 0 B; 7 C; 11 Q; 8 E; 0 Z; 18 G; 4 H;
 SQ 10 I; 14 L; 15 K; 2 M; 7 F; 5 P; 6 S; 7 T; 3 W; 6 Y; 10 V;

R42374 Length: 178 April 28, 1998 12:48 Type: P Check: 4660 ..

1 MKFTKMGWG AICSVAILA LVKANOELR ISQOGLDLIG NVEGCRDPY
 51 HCPADVLTVG IGSTEANGKN IDPKRYSDEI EIAORWAYDL RLAEQCVNRY

101 GNGKLPQGA FDAFVSITFN VCGGKQKST LFKOANOGFT POLCHOFRW
 151 IYAGGKTLG LVARRAKKA LCLGEYHD

11AA_SEQUENCE 1.0
 ID R69760 standard; Protein: 112 AA.

AC R69760;
 DT 05-OCT-1995 (first entry)
 DE PMO55973 (Met-15-125HIL-3) having an M7 substitution.
 DE Hybrid: human; murine; interleukin 3; IL-3; deletion; substitution;
 KW biological activity; side effect profile; leukotriene; release;
 KW histamine; haematopoietic cell production; leucopenia.
 OS Homo sapiens.
 OS Mus musculus.
 PN WO9500646-A.
 PD 05-JAN-1995.
 PF 17-JUN-1994; U06647.
 PR 21-JUN-1993; US-081539.
 PA (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Bradford-Goldberg SR, Easton AM, Klein BK, McKearn JP;
 PI Ollins PO;
 PI MPI: 95-052083/07.
 DR N-PSDB: Q84028.
 DR Human-murine interleukin 3 chimeric hybrid poly-peptide(s) -
 PT exhibit an improved side effects profile, e.g. reduction in
 PT stimulation of leukotriene or histamine release
 PS Claim 14; Page 126-127; 179pp; English.
 CC The sequences given in R69752-64 represent hybrid proteins containing
 CC human and murine interleukin 3 (IL-3) segments. These proteins may
 CC contain deletions at the N- and/or C-terminus, and/or deletions/
 CC substitutions within the body of the protein. These proteins have
 CC at least the biological activity of native human IL-3 and have an
 CC improved side effect profile, eg. reduction in stimulation of
 CC leukotriene release or histamine release. They may be administered
 CC in a pharmaceutical composition to stimulate haematopoietic cell
 CC production in eg. leucopenia.
 SQ Sequence 112 AA;
 SQ 8 A; 6 R; 13 N; 7 D; 0 B; 2 C; 6 Q; 8 E; 0 Z; 3 G; 3 H;
 SQ 8 I; 16 L; 5 K; 3 M; 4 F; 8 P; 4 S; 5 T; 1 W; 1 Y; 1 V;

R69760 Length: 112 April 28, 1998 12:48 Type: P Check: 1655 ..

1 MNCSSNIDEI ITHLKQPLP LLDFFNNIGE DODILMENN LRRPNLEAFNR
 51 SGQEVNNSA IESILKNLP CLPLATAPT RHPHKKDD WNEFRKXLF
 101 YKLTENAOA QO

11AA_SEQUENCE 1.0
 ID R78335 standard; Protein: 111 AA.

AC R78335;
 DT 27-JUN-1996 (first entry)
 DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
 DE Interleukin-3; mutein; colony stimulating factor; CSF; chemotherapy;
 KW radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody;
 KW synergic.
 OS Synthetic.
 PN WO9520977-A1.
 PD 10-AUG-1995.
 PF 02-FEB-1995; U01184.
 PR 04-FEB-1994; US-193373.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer CS, Bradford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins PO, Paik K;
 PI Thomas JM;
 PI MPI: 95-283609/37.
 DR Multi-lineage hematopoietic cell production - by co-administration
 PT of human interleukin-3 variants and colony stimulating factors
 PT useful in radiation treatment and chemotherapy
 PS Claim 11; Page 249; 317pp; English.
 CC R78327-R78368 are human interleukin-3 (IL-3) mutant proteins
 CC (muteins). The IL-3 muteins are used in a method for increasing the

CC prodn. of multi-lineage haematopoietic cells in a mammal. The IL-3
 CC muteins are co-administered with a colony stimulating factor
 CC selected from GM-CSF, CSF-1, G-CSF, Meg-CSF, erythropoietin, IL-1,
 CC IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13,
 CC ILF, flt3/flt2, human growth hormone, B-cell growth factor,
 CC B-cell differentiation factor, eosinophil differentiation factor
 CC and stem cell factor. Co-administration enhances therapeutic value
 CC through synergistic effects. The mutant IL-3 molecules can be used to
 CC treat conditions in which haematopoietic cells have been depleted
 CC due to e.g. chemotherapy or radiation treatment. A further use of
 CC the IL-3 variants is as antagonists of IL-3 or antigenic fragments
 CC for the prodn. of antibodies.

Sequence 111 AA:

SQ 10 A: 7 R: 12 N: 5 D: 0 B: 2 C: 8 Q: 7 E: 0 Z: 3 G: 2 H:
 SQ 9 I: 15 L: 5 K: 2 M: 4 F: 8 P: 5 S: 4 T: 1 W: 1 Y: 1 V:

R78335 Length: 111 April 28, 1998 12:48 Type: P Check: 7559 ..

1 NCSNMIDEII THIKOPPLPI IDFNINLGED QDIIMENLNR RPNLEAFNRA

51 VKSIQNASGI EAILRNLOPC LPSATAAPSR HPIIKAGDW QEFRRKLTFF

101 LKTIENNAQ Q

!!AA_SEQUENCE 1.0

ID R56050 standard; peptide; 111 AA.

AC R56050;

DT 23-JUN-1995 (first entry)

DE Human interleukin-3 mutein (PMON13350)

KW Human interleukin-3; hIL-3; mutant; mutein.

OS Synthetic.

PN WO941638-A.

PD 09-JUN-1994.

PF 22-NOV-1993: U11197.

PR 24-NOV-1992: US-981044.

PA (MONS) MONSANTO CO.

PI (SEAR) SEARLE & CO G D.

PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;

PI Easton AM, Klein BK, McKearn JP, Ollins PO, Palk K,

PI Thomas JW;

PI WPI: 94-200266/24.

DR N-PSDB: Q68098.

PT New human interleukin-3 mutant polypeptide(s) - useful for

PT treating various blood cell deficiencies, including leukaemia,

PT neutropenia and leukaemia

PS Claim 27: Page 90-91: 388pp: English.

CC The sequence is a mutein of hIL-3, comprising (15-125)hIL-3.

CC The region 15-125 comprises the substitutions: 73G, 76A, 79R, 82Q,

CC 87S, 93S, 98I, 101A, and 105Q. The muteins may have biological

CC activities similar to or better than hIL-3 and, in some cases, may

CC also have an improved side effect profile, e.g. better therapeutic

CC index. They may be used to treat e.g. leukaemia, neutropenia,

CC thrombocytopenia, (aplastic) anemia, Chediak-Higashi syndrome, SLE,

CC leukaemia, myelodysplastic syndrome, myelofibrosis; bone marrow

CC suppression or haematopoietic deficiency associated with treatment

CC with AZT or due to dialysis; or immunodeficiency resulting from viral

CC infection, exposure to radiation or cancer treatment.

CC The muteins are made by recombinant DNA techniques. Recombinant DNA

CC sequences encoding the muteins are given in GENESEQ Accession Numbers

CC Q68098 - Q68189.

SQ Sequence 111 AA:

SQ 10 A: 7 R: 12 N: 5 D: 0 B: 2 C: 8 Q: 7 E: 0 Z: 3 G: 2 H:
 SQ 9 I: 15 L: 5 K: 2 M: 4 F: 8 P: 5 S: 4 T: 1 W: 1 Y: 1 V;

R56050 Length: 111 April 28, 1998 12:48 Type: P Check: 7559 ..

1 NCSNMIDEII THIKOPPLPI IDFNINLGED QDIIMENLNR RPNLEAFNRA

51 VKSIQNASGI EAILRNLOPC LPSATAAPSR HPIIKAGDW QEFRRKLTFF

101 LKTIENNAQ Q

!!AA_SEQUENCE 1.0

ID P80381 standard; protein; 132 AA.

AC P80381;

DT 17-NOV-1990 (first entry)

DE Interleukin-3 variant.

KW Interleukin-3 variant; mutein;

KW colony stimulating factor activity.

OS Homo sapiens.

PN MO8806161-A.

PD 25-AUG-1988.

PF 18-FEB-1988: U00402.

PR 18-FEB-1987: US-016079.

PA (SCHB) Schering Biotech.

PI Otsuka T;

PI WPI: 88-249744/35.

PT Recombinant human interleukin-3 and muteins -

PT used for promoting growth and development of the hematopoietic

PT system and treating infection.

PS Disclosure; 4pp: English.

CC The human interleukin-3 variant or mutein exhibits colony

CC stimulating factor activity and comprises a glycosylated or

CC unglycosylated 10-fold substituted polypeptide where regions

CC defined by AA 10-32, 47-60, 71-86 and 95-113, inclusive, are

CC together not more than 3-fold substituted. The various

CC muteins stimulate the regeneration of blood cells and may be

CC useful in cancer therapy, treatment of certain blood

CC diseases and in the treatment of persistent infections.

CC See also P80382, P80383 and N80361.

SQ Sequence 132 AA:

SQ 10 A: 6 R: 13 N: 6 D: 0 B: 2 C: 7 Q: 7 E: 0 Z: 2 G: 3 H:
 SQ 9 I: 20 L: 7 K: 3 M: 3 F: 12 P: 6 S: 11 T: 2 W: 1 Y: 2 V;

P80381 Length: 132 April 28, 1998 12:48 Type: P Check: 4868 ..

1 PMQTPTPKT SWVNCNMID EITTHLKOPP LPLIDPNLN GEDQDILMEN

51 NLRPNLEAF NRAVKSQNA SAIESILNKL LPCUPLATAA PRRPHIRKD

101 GDWNEPRRKL TFLKLTLENA QAQTTLSLA IF

!!AA_SEQUENCE 1.0

ID P80383 standard; protein; 133 AA.

AC P80383;

DT 18-NOV-1990 (first entry)

DE Interleukin-3 variant.

KW Interleukin-3 variant; mutein;

KW colony stimulating factor activity.

OS Homo sapiens.

PN MO8806161-A.

PD 25-AUG-1988.

PF 18-FEB-1988: U00402.

PR 18-FEB-1987: US-016079.

PA (SCHB) Schering Biotech.

PI Otsuka T;

PI WPI: 88-249744/35.

PT Recombinant human interleukin-3 and muteins -

PT used for promoting growth and development of the hematopoietic

PT system and treating infection.

PS Claim 13: Page 46-47: 54pp: English.

CC The human interleukin-3 variant or mutein exhibits colony

CC stimulating factor activity and comprises a glycosylated or

CC unglycosylated 10-fold substituted polypeptide where regions

CC defined by AA 10-32, 47-60, 71-86 and 95-113, inclusive, are

CC together not more than 3-fold substituted. The various

CC muteins stimulate the regeneration of blood cells and may be

CC useful in cancer therapy, treatment of certain blood

CC diseases and in the treatment of persistent infections.

CC See also P80381, P80382 and N80361.

SQ Sequence 133 AA:

SQ 11 A: 6 R: 13 N: 6 D: 0 B: 2 C: 7 Q: 7 E: 0 Z: 2 G: 3 H:
 SQ 9 I: 20 L: 7 K: 3 M: 3 F: 12 P: 6 S: 11 T: 2 W: 1 Y: 2 V;

P80383 Length: 133 April 28, 1998 12:48 Type: P Check: 5335 ..

1 APMOTTPPK TSWWNCNNMI DEITTHKOP PLPLDPNNL NGEODITLME
51 NMLRRNLEA ENRAVKSJON ASATESILKN LPLCLPLATA APTRRPHIK
101 DGDWNEPRRK LTFYKLTLEN AQAQOTLTSL AIF

!!AA_SEQUENCE 1.0
ID R56095 standard; peptide: 113 AA.
AC R56095:
DE 26-JUN-1995 (first entry)
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN WO9412638-A.
PD 09-JUN-1994.
PF 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS) MONSANTO CO.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
PI Thomas JW;
DR N-PSDB; Q68121.
DR WPI; 94-200266/24.

PT New human interleukin-3 mutant polypeptide(s) - useful for
treating various blood cell deficiencies, including leukopenia,
neutropenia and leukaemia
PS Claim 32; Page 107-108; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
hIL-3 and, in some cases, may also have an improved side effect
profile, e.g. better therapeutic index. They may be used to treat
e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
myelofibrosis; bone marrow suppression or haematopoietic deficiency
associated with treatment with AZT or due to dialysis; or
immunodeficiency resulting from viral infection, exposure to
radiation or cancer treatment. The muteins are made by recombinant
DNA techniques. Recombinant DNA sequences encoding the muteins are
given in GENESEQ Accession Numbers Q68083 - Q68189.

SO Sequence 113 AA;
SQ 10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
11 I; 15 L; 4 K; 3 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 3 V;
R56095 Length: 113 April 28, 1998 12:48 Type: P Check: 6626 ..

1 MANCSIMIDE IHHUKRPPN PLDPNNLNS EDMDILERN LRTPLALFV
51 RAVKHEENAS GIEALIRNLQ PCUPSATAAP SRHPITIKAG DWQEFREKLT
101 FYIVTLEQAO EQQ

!!AA_SEQUENCE 1.0
ID R56105 standard; peptide: 113 AA.
AC R56105:
DE 26-JUN-1995 (first entry)
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN WO9412638-A.
PD 09-JUN-1994.
PF 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS) MONSANTO CO.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
PI Thomas JW;
DR N-PSDB; Q68131.
DR WPI; 94-200266/24.

PT New human interleukin-3 mutant polypeptide(s) - useful for
treating various blood cell deficiencies, including leukopenia,

PT neutropenia and leukaemia
PS Claim 32; Page 112; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
hIL-3 and, in some cases, may also have an improved side effect
profile, e.g. better therapeutic index. They may be used to treat
e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
myelofibrosis; bone marrow suppression or haematopoietic deficiency
associated with treatment with AZT or due to dialysis; or
immunodeficiency resulting from viral infection, exposure to
radiation or cancer treatment. The muteins are made by recombinant
DNA techniques. Recombinant DNA sequences encoding the muteins are
given in GENESEQ Accession Numbers Q68083 - Q68189.

SO Sequence 113 AA;
SQ 12 A; 6 R; 9 N; 5 D; 0 B; 2 C; 7 Q; 9 E; 0 Z; 2 G; 3 H;
9 I; 16 L; 4 K; 4 M; 3 F; 8 P; 6 S; 2 T; 1 W; 1 Y; 4 V;
R56105 Length: 113 April 28, 1998 12:48 Type: P Check: 6426 ..

1 MANCSIMIDE AIHHUKVPPA PLDSNNLNS EDMDILERN LRTPLALFV
51 RAVKHEENAS GIEALIRNLQ PCUPSATAAP SRHPITIKAG DWQEFREKLT
101 FYIVTLEQAO EQQ

!!AA_SEQUENCE 1.0
ID R56098 standard; peptide: 113 AA.
AC R56098:
DE 26-JUN-1995 (first entry)
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN WO9412638-A.
PD 09-JUN-1994.
PF 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS) MONSANTO CO.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
PI Thomas JW;
DR N-PSDB; Q68124.
DR WPI; 94-200266/24.

PT New human interleukin-3 mutant polypeptide(s) - useful for
treating various blood cell deficiencies, including leukopenia,
neutropenia and leukaemia
PS Claim 32; Page 109; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
hIL-3 and, in some cases, may also have an improved side effect
profile, e.g. better therapeutic index. They may be used to treat
e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
myelofibrosis; bone marrow suppression or haematopoietic deficiency
associated with treatment with AZT or due to dialysis; or
immunodeficiency resulting from viral infection, exposure to
radiation or cancer treatment. The muteins are made by recombinant
DNA techniques. Recombinant DNA sequences encoding the muteins are
given in GENESEQ Accession Numbers Q68083 - Q68189.

SO Sequence 113 AA;
SQ 11 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
11 I; 15 L; 4 K; 1 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;
R56098 Length: 113 April 28, 1998 12:48 Type: P Check: 6839 ..

1 MANCSIMIDE IHHUKRPPN PLDPNNLNS EDMDILERN LRTPLALFV
51 RAVKHEENAS GIEALIRNLQ PCUPSATAAP SRHPITIKAG DWQEFREKLT
101 FYIVTLEQAO EQQ

!!AA_SEQUENCE 1.0

ID R56099 standard; peptide: 113 AA.
AC R56099;
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13423).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN WO9412638-A.
PD 09-JUN-1994.
PE 22-NOV-1993: U11197.
PR 24-NOV-1993: US-981044.
PA (MONS) MONSANTO CO.
PI (SEAR) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
PI Thomas JW;
DR WPI: 94-200266/24.
DR N-PSDB: Q68125.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PS neutropenia and leukaemia
PS Claim 32: Page 109-110: 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The mutants may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENESQ Accession Numbers Q68083 - Q68189.
SQ Sequence 113 AA:
SQ 10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
SQ 12 I; 15 L; 4 K; 1 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;
R56099 Length: 113 April 28, 1998 12:48 Type: P Check: 6895 ..

1 MANSCTIIDE ITHHLKRPEN PLDDPNINNS EDDVILERN LTRPNILAV
51 RAVKHEENAS GIEAILRLNQ PCLPSATAAP SRHPITIKAG DWQEFREKLT
101 FYLVLEQAO EQQ

11AA_SEQUENCE 1.0
ID W13009 standard; protein: 560 AA.
AC W13009;
DT 21-NOV-1997 (first entry)
DE Segment of desmosomal cadherin, desmoglein Dsg2.
KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW micrometastasis; separation; enrichment; targeted delivery;
KW metastatic.
OS Homo sapiens.
PN DEL9531033-A1.
PD 27-FEB-1997.
PE 23-AUG-1995: 031033.
PR 23-AUG-1995: DE-031033.
PA (PROG) PROGEN BIOTECHNIK GMBH.
PI Franke WW, Schaefer S;
DR WPI: 97-146518/14.
PT Antibody reactive with part of desmosomal cadherin - exposed on
PT surface of epithelial or carcinoma cells, not bound to desmosomes,
PS useful for diagnosis and treatment of carcinoma micrometastases
PS Claim 7: Page 5; 8pp; German.
CC The present sequence is a segment of the desmosomal cadherin (DC),
CC desmoglein Dsg2, which is exposed on the surface of epithelial or
CC carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC directed against epitopes of the present sequence can be used to
CC diagnose, i.e. to detect carcinoma cells, especially
CC micrometastases, not bound to desmosomes, to separate, enrich or
CC detect living or fixed carcinoma cells by cell sorting methods and

CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC target cells. The Ab provides rapid and reliable detection of
CC metastatic carcinoma, and detects parts of DC that are not
CC accessible in desmosome bound cells, as in normal tissue or
CC carcinomas.
SQ Sequence 560 AA:
SQ 31 A; 19 R; 33 N; 39 D; 0 B; 5 C; 17 Q; 50 E; 0 Z; 32 G; 11 H;
SQ 43 I; 38 L; 38 K; 5 M; 22 F; 26 P; 39 S; 38 T; 4 W; 18 Y; 52 V;
W13009 Length: 560 April 28, 1998 12:48 Type: P Check: 3849 ..

1 AMITAPVALR EGEDLSKKNP IAKIHSDLAE ENGLKITRYX TSGKITEPFP
51 GIVFENKDG ELNVTSLDR EETPFLLTG YALDAGNNV EXPLERIKV
101 LDINDNEPVF TQDVVGSVE ELSAHTLVN KINATDADEP MTLNKSISR
151 IYSLERAYPP VFYLNKDTGE IYTSVTLDR EEHSSTYLV EARDNGEYV
201 DKPVKAQVQ IRLDVNDNI PVVENKVLGE MVENQVNE VTRIVFPAD
251 EIGSDMWLAN PTFASGNEG YFHIEDAQT NEGIVTLKE VDYEEKNID
301 FSVYANKAA FSDSRSRYK PTPIPIKVY KNYKEGIRK SSVISIVSE
351 SMDRSSKGI IGNFAPED TGLPAHARYV KLEDRDNWIS VDSVSEIKL
401 AKIPDEFNR VONGTYTYKI VAISEDYPRK TITGYLVIN EDINDCPTL
451 IEFVQITCD AEYVAVTAD LDGHPNSGPF SFSVIDKPG MAEKKIRAO
501 ESTSVLQGS EKKLGRSEIQ FLISDNGFS CPKQVLTIL VCVLHGSGC
551 REAQDSYVG

11AA_SEQUENCE 1.0
ID R56099 standard; peptide: 113 AA.
AC R56099;
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13409).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN WO9412638-A.
PD 09-JUN-1994.
PE 22-NOV-1993: U11197.
PR 24-NOV-1992: US-981044.
PA (MONS) MONSANTO CO.
PI (SEAR) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
PI Thomas JW;
DR WPI: 94-200266/24.
DR N-PSDB: Q68122.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PS neutropenia and leukaemia
PS Claim 32: Page 108; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The mutants may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENESQ Accession Numbers Q68083 - Q68189.
SQ Sequence 113 AA:
SQ 10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
SQ 10 I; 16 L; 4 K; 3 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 3 V;


```

101 FYLVLEQAO EQQ
!!AA_SEQUENCE 1.0
ID R56137 standard; peptide; 113 AA.
AC R56137;
DE 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13440).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN MO9412638-A.
PD 09-JUN-1994.
PE 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Bratford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
PI Thomas JW;
PI WPI: 94-200266/24.
DR N-PSDB; Q68170.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 32; Page 130-131; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENSEQ Accession Numbers Q68083 - Q68189.
CC Sequence 113 AA:
SQ 12 A; 6 R; 8 N; 4 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
SQ 10 I; 14 L; 4 K; 3 M; 3 F; 8 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
SQ 10 I; 14 L; 4 K; 3 M; 3 F; 8 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
R56137 Length: 113 April 28, 1998 12:48 Type: P Check: 7231 ..

1 MANCSIMIDE AIHRLKIPPN PSIDSANLNS EDVSIIMERN LRTPNLAFV
51 RAVKLENAS GIEALIRNLQ PCLPSTAP SRHPIIKAG DMQFRREKLT
101 FYLVLEQAO EQQ

!!AA_SEQUENCE 1.0
ID R78368 standard; protein; 113 AA.
AC R78368;
DE 27-JUN-1996 (first entry)
DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
KW Interleukin-3; mutein; colony stimulating factor; CSF; chemotherapy;
KW radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody;
OS Synthetic.
PN MO9520977-A1.
PD 10-AUG-1995.
PE 02-FEB-1995; U01184.
PR 04-FEB-1994; US-193373.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer CS, Bratford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
PI Thomas JW;
PI WPI: 95-283609/37.
PT Multi-lineage haematopoietic cell production - by co-administration
PT of human interleukin-3 variants and colony stimulating factors
PT useful in radiation treatment and chemotherapy
PS Claim 12; Page 238; 317pp; English.
CC R78327-R78368 are human interleukin-3 (IL-3) mutant proteins
CC (muteins). The IL-3 muteins are used in a method for increasing the
CC prodn. of multi-lineage haematopoietic cells in a mammal. The IL-3

```

```

CC muteins are co-administered with a colony stimulating factor
CC selected from GM-CSF, CSF-1, G-CSF, Meg-CSF, erythropoietin, IL-1,
CC IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13,
CC LIF, flt3/flt2, human growth hormone, B-cell growth factor,
CC B-cell differentiation factor, eosinophil differentiation factor
CC and stem cell factor. Co-administration enhances therapeutic value
CC through synergistic effects. The mutant IL-3 molecules can be used to
CC treat conditions in which haematopoietic cells have been depleted
CC due to e.g. chemotherapy or radiation treatment. A further use of
CC the IL-3 variants is as antagonists of IL-3 or antigenic fragments
CC for the prodn. of antibodies.
CC Sequence 113 AA:
SQ 11 A; 6 R; 9 N; 4 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
SQ 9 I; 15 L; 4 K; 3 M; 3 F; 8 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
SQ 9 I; 15 L; 4 K; 3 M; 3 F; 8 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
R78368 Length: 113 April 28, 1998 12:48 Type: P Check: 7382 ..

1 MANCSIMIDE LIHRLKIPPN PSIDSANLNS EDVSIIMERN LRTPNLAFV
51 RAVKLENAS GIEALIRNLQ PCLPSTAP SRHPIIKAG DMQFRREKLT
101 FYLVLEQAO EQQ

!!AA_SEQUENCE 1.0
ID R56120 standard; peptide; 113 AA.
AC R56120;
DE 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13383).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN MO9412638-A.
PD 09-JUN-1994.
PE 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Bratford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
PI Thomas JW;
PI WPI: 94-200266/24.
DR N-PSDB; Q68146.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Disclosure; Page 119; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENSEQ Accession Numbers Q68083 - Q68189.
CC Sequence 113 AA:
SQ 11 A; 7 R; 9 N; 6 D; 0 B; 2 C; 7 Q; 9 E; 0 Z; 2 G; 3 H;
SQ 9 I; 15 L; 4 K; 3 M; 3 F; 9 P; 6 S; 2 T; 1 W; 1 Y; 4 V;
SQ 9 I; 15 L; 4 K; 3 M; 3 F; 9 P; 6 S; 2 T; 1 W; 1 Y; 4 V;
R56120 Length: 113 April 28, 1998 12:48 Type: P Check: 7137 ..

1 MANCSIMIDE AIHRLKIPPN PLIDPNLND EDVSIIMERN LRLPNLESFV
51 RAVKLENAS GIEALIRNLQ PCLPSTAP SRHPIIKAG DMQFRREKLT
101 FYLVLEQAO EQQ

!!AA_SEQUENCE 1.0
ID R56067 standard; peptide; 113 AA.
AC R56067;
DE 22-JUN-1995 (first entry)

```

DE Human interleukin-3 mutein (pMon13288).
 KW Human interleukin-3; hIL-3; mutant; mutein.
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH,
 PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
 PI Thomas JW;
 DR WPI: 94-200266/24.
 DR N-PSDB: Q68086.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 27; Page 99; 388pp; English.
 CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
 CC The region 15-125 comprises the substitutions: 181, 25H, 29R,
 CC 32N, 37P, 42S, 45W, 51R, 55T, 59L, 62V, 67H, 69E, 73G, 76A, 79R,
 CC 82Q, 87S, 93S, 98T, 101A, 105Q, 109P, 116V, 120Q, 123E.
 CC The muteins may have biological activities similar to or
 CC better than hIL-3 and, in some cases, may also have an improved side
 CC effect profile, e.g. better therapeutic index. They may be used to
 CC treat e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to radiation
 CC or cancer treatment.
 CC The muteins are made by recombinant DNA techniques. Recombinant DNA
 CC sequences encoding the muteins are given in GENESEQ Accession Numbers
 CC Q68083 - Q68189.
 SO Sequence 113 AA:
 SO 10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
 SO 10 I; 15 L; 4 K; 4 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 3 V;
 RS6067 Length: 113 April 28, 1998 12:48 Type: P Check: 6774 ..

1 MANCSIMIDE ITHHKRPPN PLIDRNNLNS EDMOIMERN LRTPNLLAFV
 51 RAVKHELENAS GIEALIRNLQ PCLPSATAP SRHPIIKAG DMQEFREKLT
 101 FYLVTEQAO EQQ

11AA_SEQUENCE 1.0
 ID R56083 standard; peptide: 113 AA.
 AC R56083;
 DE 29-JUN-1995 (first entry)
 KW Human interleukin-3 mutein (pMon13402).
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH,
 PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
 PI Thomas JW;
 DR WPI: 94-200266/24.
 DR N-PSDB: Q68169.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 27; Page 130; 388pp; English.
 CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
 CC The muteins may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The muteins are made by recombinant
 CC DNA techniques. Recombinant DNA sequences encoding the muteins are
 CC given in GENESEQ Accession Numbers Q68083 - Q68189.
 SO Sequence 113 AA:
 SO 12 A; 6 R; 8 N; 4 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
 SO 9 I; 14 L; 4 K; 3 M; 3 F; 9 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
 RS6136 Length: 113 April 28, 1998 12:48 Type: P Check: 7287 ..

1 MANCSIMIDE AIHHKIRPPN PSLDSANLNS EDVSIIMERN LRTPNLLAFV
 51 RAVKHELENAS GIEALIRNLQ PCLPSATAP SRHPIIKAG DMQEFREKLT
 101 FYLVTEQAO EQQ

11AA_SEQUENCE 1.0
 ID R56136 standard; peptide: 113 AA.
 AC R56136;
 DE 29-JUN-1995 (first entry)
 KW Human interleukin-3 mutein (pMon13377).
 OS Synthetic.
 PN W09412638-A.

CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The muteins are made by recombinant
 CC DNA techniques. Recombinant DNA sequences encoding the muteins are
 CC given in GENESEQ Accession Numbers Q68083 - Q68189.
 SO Sequence 113 AA:
 SO 11 A; 6 R; 8 N; 4 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
 SO 10 I; 15 L; 4 K; 3 M; 3 F; 9 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
 RS6083 Length: 113 April 28, 1998 12:48 Type: P Check: 7352 ..

1 MANCSIMIDE LIHHKIRPPN PSLDSANLNS EDVSIIMERN LRTPNLLAFV
 51 RAVKHELENAS GIEALIRNLQ PCLPSATAP SRHPIIKAG DMQEFREKLT
 101 FYLVTEQAO EQQ

11AA_SEQUENCE 1.0
 ID R56136 standard; peptide: 113 AA.
 AC R56136;
 DE 29-JUN-1995 (first entry)
 KW Human interleukin-3 mutein (pMon13400).
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH,
 PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
 PI Thomas JW;
 DR WPI: 94-200266/24.
 DR N-PSDB: Q68168.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 32; Page 129-130; 388pp; English.
 CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
 CC The muteins may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The muteins are made by recombinant
 CC DNA techniques. Recombinant DNA sequences encoding the muteins are
 CC given in GENESEQ Accession Numbers Q68083 - Q68189.
 SO Sequence 113 AA:
 SO 12 A; 6 R; 8 N; 4 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
 SO 9 I; 14 L; 4 K; 3 M; 3 F; 9 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
 RS6136 Length: 113 April 28, 1998 12:48 Type: P Check: 7287 ..

1 MANCSIMIDE AIHHKIRPPN PSLDSANLNS EDVSIIMERN LRTPNLLAFV
 51 RAVKHELENAS GIEALIRNLQ PCLPSATAP SRHPIIKAG DMQEFREKLT
 101 FYLVTEQAO EQQ

11AA_SEQUENCE 1.0
 ID R56115 standard; peptide: 113 AA.
 AC R56115;
 DE 26-JUN-1995 (first entry)
 KW Human interleukin-3 mutein (pMon13377).
 OS Synthetic.
 PN W09412638-A.

```

PD 09-JUN-1994.
PF 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKeern JP, Olin PO, Paik K;
PI Thomas JW;
DR WPI; 94-200266/24.
DR N-PSDB; 068141.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 32; Page 116-117; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENESEQ Accession Numbers 068083 - 068189.
SQ Sequence 113 AA;
SO 11 A; 7 R; 9 N; 6 D; 0 B; 2 C; 7 Q; 9 E; 0 Z; 2 G; 3 H;
SO 9 I; 15 L; 4 K; 3 M; 3 F; 9 P; 6 S; 2 T; 1 W; 1 Y; 4 V;
R56115 Length: 113 April 28, 1998 12:48 Type: P Check: 7137 ..

1 MANCSIMIDE AIHHLKRPFA PLDDPNLND EDVSLMDRN LRLPLNESVY
51 RAVKLEENAS GIEALIRNLQ PCLPSATRAP SRAPITIKAG DMQERRKIQ
101 FYLVTEQAO EQQ

11AA_SEQUENCE 1.0
ID R10727 standard; Protein: 145 AA.
AC R10727;
DE 22-APR-1991 (first entry)
DE (Glu108Asp109Glu110)IL-3.
KW Interleukin-3; mutant; antagonist; cytopenia; immunodeficiency;
KW immunosuppression; charge reversal.
FH Key Location/Qualifiers
FT Peptide 1..12
FT /label= signal sequence
FT Mutation 120..122
FT /label= Arg-Arg-Lys -> Glu-Asp-Glu
PN EP-413383-A.
PD 20-FEB-1991.
PF 02-AUG-1990; 202117.
PF 14-AUG-1989; EP-202082.
PR 14-SEP-1989; EP-202331.
PR 02-AUG-1990; EP-202117.
PA (KONN ) GIST-BROCADES NV.
PI Dorsers ICI, Vanleen RW;
PI WPI; 91-052685/08.
DR New interleukin-3 mutants - having deletions covering complete
PT coding sequence while retaining biological activity, have
PT antagonistic effect.
PS Claim 6; Page 18; 26pp; English.
CC The mutant was prepd. by recombinant DNA techniques using mutagenic
CC primers to change the codons of cDNA encoding IL-3. The peptide
CC pref. also has a deletion at the N-terminus (Aas 1-14) and/or the
CC C-terminus (Aas 116-133,120-130 and 130-133). The peptide has
CC antagonistic activity and can be used for therapeutic and diagnostic
CC purposes.
CC See also R10711-R10731 and R11051-R11055.
SQ Sequence 145 AA;
SO 11 A; 5 R; 14 N; 8 D; 0 B; 2 C; 7 Q; 9 E; 0 Z; 3 G; 3 H;
SO 10 I; 20 L; 6 K; 5 M; 5 F; 10 P; 8 S; 13 T; 2 W; 1 Y; 3 V;

```

```

R10727 Length: 145 April 28, 1998 12:48 Type: P Check: 7418 ..

1 MMINRSRS VDAPIQTPT LKTSWNCN MDEITTHK QPPLPLDFN
51 NINGEDDIL MENILRRPNL EAFNRVAKSL QNASALESTL KNLPLCPLPA
101 TAAPTRHPFH IKDGDWNEFE DELTFYIKTL ENAQAOQTTL SLAIF

11AA_SEQUENCE 1.0
ID R14839 standard; Protein: 464 AA.
AC R14839;
DE 30-JAN-1992 (first entry)
DE Protein deduced from human Glioblastoma cell cDNA.
KW RAS; oncogene; cancer; cAMP; p2165.
OS Homo sapiens.
PN WO9116457-A.
PD 31-OCT-1991.
PF 19-APR-1991; U02714.
PR 20-APR-1990; US-511715.
PA (COLD-) COLD SPRING HARBOR.
PI Wigler MH, Colicelli JJ;
PI WPI; 91-339841/46.
DR N-PSDB; Q14627.
PT Complementary screening for genes and prods. - e.g. RAS protein
PT and cAMP, that modify, complement or suppress genetic defect and
PT correct associated phenotypic alteration
PS Example 1b; Page 77; 169pp; English.
CC In the specification the sequence from which this amino acid
CC sequence was deduced is given the SEQ ID NO. 15 and
CC is described as the cDNA insert of plasmid p2165. The insert
CC sequence shows no significant homology to previously isolated
CC genes. Plasmid p2165 in E.coli (ATCC 68598) is described as
CC containing a human glioblastoma cell cDNA insert encoding a
CC RAS-related polypeptide. N.B. Many of the references in the text of
CC the specification to particular sequences are incorrect, therefore
CC caution is advised in accepting that this sequence is indeed p2165 !
SO Sequence 464 AA;
SO 18 A; 22 R; 14 N; 28 D; 0 B; 9 C; 27 Q; 38 E; 0 Z; 17 G; 8 H;
SO 20 I; 47 L; 34 K; 20 M; 22 F; 22 P; 39 S; 27 T; 3 W; 19 Y; 30 V;
R14839 Length: 464 April 28, 1998 12:48 Type: P Check: 4420 ..

1 MSISRSSSDS LEFDRSMPLF GYEADNSSL EYEBESDDE TMAPRIRKSK
51 KRSSSFVLPK LVKSOLQKVS GVSSSEMTPE KMWRRMAEL SRDKCTYFGC
101 LVQDYVSFLQ ENKECHVST DMLQITRQFM TQVKNYLSQS SELDPPIESL
151 IREDQIDVTL EKAMHKCLIK PLKGAVEAML KQFNHAGSM KQKKNQLQV
201 RQRNPQELGV FAPTRDFVDV EKIKVFMFM QKMSPEKKV MLLLRVCKLI
251 YTVMENNSGR MYGADDELPLV LTFVIAQCDW LELDTEIEYM MELDPSILH
301 GEGGYLLISA YGALSILKNF QEQOARILS SETROTIRQM HKRRTNTTI
351 PSYDQFQNTL RVAFQVNSG CTGKTLIVRP YTTEDVCOI CAEKFKVADP
401 EYSLFLFVN ETWOOLAEDT YPOKIRAEHL SRPQHIHF YKRIKINDPY
451 GILFQNGEED LTTS

11AA_SEQUENCE 1.0
ID R78367 standard; Protein: 125 AA.
AC R78367;
DE 27-JUN-1996 (first entry)
DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
DE Interleukin-3; mutein; colony stimulating factor; CSF; chemotherapy;
KW radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody;
KW synergy.
OS Synthetic.
PN WO9520977-A1.

```

PD 10-AUG-1995.
 PR 02-FEB-1995; U01184.
 PR 04-FEB-1994; US-193373.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, Mckearn JP, Olins PO, Palk K;
 PI Thomas JM;
 DR WPI; 95-283609/37.
 PT Multi-lineage hematopoietic cell production - by co-administration
 of human interleukin-3 variants and colony stimulating factors
 useful in radiation treatment and chemotherapy
 PS Claim 11; Page 258; 317pp; English.
 CC R78327-R78358 are human Interleukin-3 (IL-3) mutant proteins
 (mutelns). The IL-3 mutelns are used in a method for increasing the
 prodn. of multi-lineage haematopoietic cells in a mammal. The IL-3
 CC mutelns are co-administered with a colony stimulating factor
 CC selected from GM-CSF, CSF-1, G-CSF, Meg-CSF, erythropoietin, IL-1,
 CC IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13,
 CC ILF, IL13/IL12, human growth hormone, B-cell growth factor,
 CC B-cell differentiation factor, eosinophil differentiation factor
 CC and stem cell factor. Co-administration enhances therapeutic value
 CC through synergistic effects. The mutant IL-3 molecules can be used to
 CC treat conditions in which haematopoietic cells have been depleted
 CC due to e.g. chemotherapy or radiation treatment. A further use of
 CC the IL-3 variants is as antagonists of IL-3 or antigenic fragments
 CC for the prodn. of antibodies.
 SQ Sequence 125 AA;
 SQ 10 A; 7 R; 9 N; 10 D; 0 B; 2 C; 6 Q; 10 E; 0 Z; 2 G; 4 H;
 SQ 10 I; 15 L; 6 K; 4 M; 3 F; 10 P; 5 S; 5 T; 1 W; 3 Y; 3 V;
 R78367 Length: 125 April 28, 1998 12:48 Type: P Check: 5328 ..

1 MAYPETYKD DDDKNCISIM I DEIHHKRP PNPLDPNNL NSEMDITLME
 51 RNLTPTNLLA FVRAVKHLEN NASGIEAILRN LQPCLPSSAT APSRHPITIK
 101 AGDMQEFREK LTFYLVLTLEQ AQOEOQ

11AA_SEQUENCE 1.0
 ID R56082 standard; peptide; 126 AA.
 AC R56082;
 DT 30-JUN-1995 (first entry)
 DE Human interleukin-3 muteln (pMon13390).
 KW Human interleukin-3; hIL-3; mutant; muteln.
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, Mckearn JP, Olins PO, Palk K;
 PI Thomas JM;
 DR WPI; 94-200266/24.
 DR N-PSDB; Q68183.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 27; Page 129; 388pp; English.
 CC The sequence is a muteln of hIL-3, comprising Met-Ala-Tyr-Pro-Glu-
 CC Thr-Asp-Tyr-Lys-Asp-Asp-Lys-Ala-(15-15)hIL-3.
 CC The mutelns may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The mutelns are made by recombinant
 CC DNA techniques. Recombinant DNA sequences encoding the mutelns are
 CC given in GENESDB Accession Numbers Q68083 - Q68189. NOTE: The

CC protein sequence given in the specification lacks the alanine
 CC residue at position 15 although the text associated with the
 CC corresponding nucleotide sequence (Q68183, Page 164) suggests that
 CC this alanine residue should be present. The polypeptide translated
 CC from the nucleotide sequence does decode to give the alanine
 CC residue.
 SQ Sequence 126 AA;
 SQ 11 A; 7 R; 9 N; 10 D; 0 B; 2 C; 6 Q; 10 E; 0 Z; 2 G; 4 H;
 SQ 10 I; 15 L; 6 K; 4 M; 3 F; 10 P; 5 S; 5 T; 1 W; 3 Y; 3 V;
 R56082 Length: 126 April 28, 1998 12:48 Type: P Check: 5172 ..

1 MAYPETYKD DDDKNCISIM IDEIHHKRP PNPLDPNN NSEMDITLME
 51 ERNLTPTNLL AFVRAVKHLE NASGIEAILR NLQPCLPSSAT AASRHPITIK
 101 KAGDMQEFRE KTFYLVLTLE QAOEOQ

11AA_SEQUENCE 1.0
 ID R72583 standard; Peptide; 15 AA.
 AC R72583;
 DT 29-SEP-1995 (first entry)
 DE EMAPRI active peptide.
 KW EMAPRI; endothelial monocyte activating polypeptide II; chemotaxis;
 KW Inflammation; tissue factor; tumor; cancer; therapy; metha;
 KW sarcoma.
 OS Synthetic.
 PN W09509180-A.
 PD 06-APR-1995.
 PF 29-SEP-1994; U11085.
 PR 29-SEP-1993; US-129456.
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 PI Clause M, Kao J, Kayton M, Labutti SK, Stern DM;
 DR WPI; 95-147389/19.
 PT New endothelial monocyte activating polypeptide II - induces
 PT chemotaxis, inflammation and tissue factor, useful for treating
 PT tumours, also related antibodies, DNA and active fragments
 PS Claim 36; Page 129; 180pp; English.
 CC EMAPRI protein was purified from conditioned medium of mouse
 CC metha sarcoma cells. The protein includes the N-terminal
 CC sequence given in R72578. Peptides (R72580-87) based on this
 CC sequence were able to induce chemotaxis in polymorphonuclear
 CC leukocytes and mononuclear phagocytes.
 SQ Sequence 15 AA;
 SQ 2 A; 3 R; 0 N; 1 D; 0 B; 0 C; 0 Q; 0 E; 0 Z; 1 G; 0 H;
 SQ 2 I; 2 L; 1 K; 0 M; 0 F; 0 P; 1 S; 1 T; 0 W; 0 Y; 1 V;
 R72583 Length: 15 April 28, 1998 12:48 Type: P Check: 9156 ..

1 ASRLDLRIGR IYVAK
 11AA_SEQUENCE 1.0
 ID R56093 standard; peptide; 113 AA.
 AC R56093;
 DT 26-JUN-1995 (first entry)
 DE Human interleukin-3 muteln (pMon13405).
 KW Human interleukin-3; hIL-3; mutant; muteln.
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, Mckearn JP, Olins PO, Palk K;
 PI Thomas JM;
 DR WPI; 94-200266/24.
 DR N-PSDB; Q68119.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 32; Page 107; 388pp; English.

```

CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENSEQ Accession Numbers Q68083 - Q68189.
SQ Sequence 113 AA:
SQ 11 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
SQ 10 I; 15 L; 4 K; 2 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;
R56093 Length: 113 April 28, 1998 12:48 Type: P Check: 6987 ..

1 MANCSIAIDE IHHKRPNN PLDPNNLNS EDVDLMERN LRTPNLAFV
51 RAVKHEUNAS GIEALRNMQ PCIPSATAAP SRHPIIKAG DWQPEREKL
101 FYLVTLQAO EQQ

!!AA_SEQUENCE 1.0
ID R56068 standard; peptide: 113 AA.
AC R56068:
DT 22-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13289).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN MO9412638-A.
PD 09-JUN-1994.
PE 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
PI Thomas JW;
PI WPI: 94-200266/24.
DR N-PSDB: Q68089.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 27, Page 99; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The region 15-125 comprises the substitutions: 181, 25H, 29V,
CC 32A, 37S, 42S, 45M, 51R, 55L, 59L, 62V, 67N, 69E, 73G, 76A, 79R,
CC 82Q, 87S, 93S, 98I, 101A, 105Q, 109E, 116V, 120Q, 123E.
CC The muteins may have biological activities similar to or
CC better than hIL-3 and, in some cases, may also have an improved side
CC effect profile, e.g. better therapeutic index. They may be used to
CC treat e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to radiation
CC or cancer treatment.
CC The muteins are made by recombinant DNA techniques. Recombinant DNA
CC sequences encoding the muteins are given in GENSEQ Accession Numbers
CC Q68083 - Q68189.
SQ Sequence 113 AA:
SQ 11 A; 6 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 3 H;
SQ 10 I; 16 L; 4 K; 4 M; 3 F; 8 P; 6 S; 3 T; 1 W; 1 Y; 4 V;
R56068 Length: 113 April 28, 1998 12:48 Type: P Check: 6643 ..

1 MANCSIMIDE IHHKVPRA PLDSNNLNS EDMDLERN LRTPLAFV
51 RAVKHEUNAS GIEALRNMQ PCIPSATAAP SRHPIIKAG DWQPEREKL
101 FYLVTLQAO EQQ

```

```

!!AA_SEQUENCE 1.0
ID R56092 standard; peptide: 113 AA.
AC R56092:
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13407).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN MO9412638-A.
PD 09-JUN-1994.
PE 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
PI Thomas JW;
PI WPI: 94-200266/24.
DR N-PSDB: Q68118.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 32, Page 106; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENSEQ Accession Numbers Q68083 - Q68189.
SQ Sequence 113 AA:
SQ 10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
SQ 10 I; 15 L; 4 K; 3 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;
R56092 Length: 113 April 28, 1998 12:48 Type: P Check: 7071 ..

1 MANCSIMIDE IHHKRPNN PLDPNNLNS EDVDLMERN LRTPNLAFV
51 RAVKHEUNAS GIEALRNMQ PCIPSATAAP SRHPIIKAG DWQPEREKL
101 FYLVTLQAO EQQ

!!AA_SEQUENCE 1.0
ID R56101 standard; peptide: 113 AA.
AC R56101:
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13425).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN MO9412638-A.
PD 09-JUN-1994.
PE 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
PI Thomas JW;
PI WPI: 94-200266/24.
DR N-PSDB: Q68127.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 32, Page 110; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat

```


CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The mutants are made by recombinant
 CC DNA techniques. Recombinant DNA sequences encoding the mutants are
 CC given in GENESQ Accession Numbers Q68083 - Q68189.

SO Sequence 113 AA:
 SO 10 A: 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
 SO 11 I; 16 L; 4 K; 1 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;
 SO 10 I; 16 L; 4 K; 1 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;

R56101 Length: 113 April 28, 1998 12:48 Type: P Check: 7006 ..

1 MANCISIIDE IHHKRPNN PLDPNNLNS EDVDILLERN LRTPNLLAFV
 51 RAVKHLLENAS GIEALIRNLQ PCLPSATAP SRHPITIKAG DMQEFREKLT
 101 FYLVTLQAO EQQ

11AA_SEQUENCE 1.0
 ID R78353 standard; protein; 113 AA.
 AC R78353;
 DT 27-JUN-1996 (first entry)
 DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
 KW Interleukin-3; mutin; colony stimulating factor; CSF; chemotherapy;
 KW radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody;
 KW synergic.
 OS Synthetic.
 PN W09520977-A1.
 PD 10-AUG-1995.
 PE 02-FEB-1995; U01184.
 PR 04-FEB-1994; US-193373.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins FO, Palk K;
 PI Thomas JW;
 DR WPI: 95-283609/37.
 PT Multi-lineage hematopoietic cell production - by co-administration
 PT of human interleukin-3 variants and colony stimulating factors
 PT useful in radiation treatment and chemotherapy

PS Claim 11; Page 254; 317pp; English.
 CC R7837-R7838 are human interleukin-3 (IL-3) mutant proteins
 CC (mutins). The IL-3 mutins are used in a method for increasing the
 CC prodn. of multi-lineage haematopoietic cells in a mammal. The IL-3
 CC mutins are co-administered with a colony stimulating factor
 CC selected from GM-CSF, CSF-1, G-CSF, Meg-CSF, erythropoietin, IL-1,
 CC IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13,
 CC LIF, flt3/flt2, human growth hormone, B-cell growth factor,
 CC B-cell differentiation factor, eosinophil differentiation factor
 CC and stem cell factor. Co-administration enhances therapeutic value
 CC through synergistic effects. The mutant IL-3 molecules can be used to
 CC treat conditions in which haematopoietic cells have been depleted
 CC due to e.g. chemotherapy or radiation treatment. A further use of
 CC the IL-3 variants is as antagonists of IL-3 or antigenic fragments
 CC for the prodn. of antibodies.

SO Sequence 113 AA:
 SO 11 A: 6 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 3 H;
 SO 10 I; 16 L; 4 K; 4 M; 3 F; 8 P; 6 S; 3 T; 1 W; 1 Y; 4 V;
 SO 10 I; 16 L; 4 K; 4 M; 3 F; 8 P; 6 S; 3 T; 1 W; 1 Y; 4 V;

R78353 Length: 113 April 28, 1998 12:48 Type: P Check: 6643 ..

1 MANCISIMIDE IHHKVPRA PLLDSNNLNS EDMDILERN LRLPNLLAFV
 51 RAVKHLLENAS GIEALIRNLQ PCLPSATAP SRHPITIKAG DMQEFREKLT
 101 FYLVTLQAO EQQ

11AA_SEQUENCE 1.0
 ID R56100 standard; peptide; 113 AA.
 AC R56100;
 DT 26-JUN-1995 (first entry)
 DE Human interleukin-3 mutin (pMon13424).

KW Human interleukin-3; hIL-3; mutant; mutin.
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PE 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins FO, Palk K;
 PI Thomas JW;
 DR WPI: 94-200266/24.
 DR N-PSDB; Q68126.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 32; Page 110; 388pp; English.
 CC The sequence is a mutin of hIL-3, comprising Met-Ala-(15-125)hIL-3.
 CC The mutants may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The mutants are made by recombinant

11AA_SEQUENCE 1.0
 ID R56104 standard; peptide; 113 AA.
 AC R56104;
 DT 26-JUN-1995 (first entry)
 DE Human interleukin-3 mutin (pMon13368).
 KW Human interleukin-3; hIL-3; mutant; mutin.
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PE 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins FO, Palk K;
 PI Thomas JW;
 DR WPI: 94-200266/24.
 DR N-PSDB; Q68130.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 32; Page 111-112; 388pp; English.
 CC The sequence is a mutin of hIL-3, comprising Met-Ala-(15-125)hIL-3.
 CC The mutants may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The mutants are made by recombinant

1 MANCISIIDE IHHKRPNN PLDPNNLNS EDVDILLERN LRTPNLLAFV
 51 RAVKHLLENAS GIEALIRNLQ PCLPSATAP SRHPITIKAG DMQEFREKLT
 101 FYLVTLQAO EQQ

R56100 Length: 113 April 28, 1998 12:48 Type: P Check: 6550 ..

CC DNA techniques. Recombinant DNA sequences encoding the muteins are
 given in GENSEQ Accession Numbers Q68083 - Q68189.
 SO Sequence 113 AA;
 SO 12 A; 6 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 3 H;
 SO 9 I; 16 L; 4 K; 4 M; 3 F; 8 P; 6 S; 3 T; 1 W; 1 Y; 4 V;

R56104 Length: 113 April 28, 1998 12:48 Type: P Check: 6555 ..

1 MANCISIMIDE AIHHLKVPFA PLIDSNLNS EDMILMERN LRLPNLAFV
 51 RAVKLENAS GIEALIRNLQ PCLPSATAAP SRHPITIKAG DMOEFREKLT
 101 FYLVLEQAO EQQ

11AA_SEQUENCE 1.0
 ID R56094 standard; peptide; 113 AA.
 AC R56094;

DT 26-JUN-1995 (first entry)
 DE Human interleukin-3 mutein (pMon13415).
 KW Human interleukin-3; hIL-3; mutant; mutein.
 OS Synthetic.
 PN WO9412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.

PA (MONS) MONSANTO CO.
 (SEAR) SEARLE & CO G D.
 PI Abrams M, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
 PI Thomas JW;
 DR WPI: 94-200266/24.
 DR N-PSDB: Q68120.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia

PS Claim 32: Page 107; 38pp; English.
 CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
 CC The muteins may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The muteins are made by recombinant
 CC DNA techniques. Recombinant DNA sequences encoding the muteins are
 CC given in GENSEQ Accession Numbers Q68083 - Q68189.
 SO Sequence 113 AA;
 SO 10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
 SO 11 I; 15 L; 4 K; 2 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;

R56094 Length: 113 April 28, 1998 12:48 Type: P Check: 7043 ..

1 MANCISIIDE IHHHLKRPN PLIDPNLNS EDVIMLERN LRLPNLAFV
 51 RAVKLENAS GIEALIRNLQ PCLPSATAAP SRHPITIKAG DMOEFREKLT
 101 FYLVLEQAO EQQ

11AA_SEQUENCE 1.0
 ID R22820 standard; Protein; 134 AA.
 AC R22820;

DT 04-SEP-1992 (first entry)
 DE K64, P83, M89, K113, E122, M127, E132, S134 hIL-36 derivative.
 KW Human, interleukin-3; bone marrow transplant; graft; platelet.
 OS Homo sapiens.
 PN J04063595-A.
 PD 28-FEB-1992.
 PF 19-OCT-1990; JP-087468.
 PR 03-APR-1990; JP-087468.
 PR 19-OCT-1990; JP-279108.

PA (KIRI) KIRIN BREWERY KK.

DR WPI: 92-120155/15.
 DR N-PSDB: Q22508.

PT Human interleukin 3 deriv. and its prepn. - for supplementing
 PT bone marrow transplantation and increasing platelet count
 PS Disclosure; Fig 8; 21pp; Japanese.
 CC This derivative of human IL-3 has amino acids Arg, Leu, Leu, Thr,
 CC Ala, Thr, Ala and Phe at positions 64, 83, 88, 113, 122, 127, 132
 CC and 134, respectively, substituted by Lys, Pro, Met, Lys, Glu, Met,
 CC Glu and Ser. The derivative has a higher activity than native
 CC IL-3. See R22813-4, Q22503-Q22510 and Q22513.
 SO Sequence 134 AA;
 SO 9 A; 5 R; 13 N; 6 D; 0 B; 2 C; 7 Q; 9 E; 0 Z; 2 G; 3 H;
 SO 9 I; 18 L; 9 K; 6 M; 4 F; 10 P; 8 S; 9 T; 2 W; 1 Y; 2 V;

R22820 Length: 134 April 28, 1998 12:48 Type: P Check: 6016 ..

1 MAPMTQTSL KISWVNSNM IDEITHLKO PPLPLDPNN LNEGDDILM
 51 ENNLRRPMLE AFNKAVALQ NASALSIK NLPCLPMAT AAPTRHPDHI
 101 KDGDMNEFR KLFYKLKLE NEQAQMTLS LEIS

11AA_SEQUENCE 1.0
 ID R56139 standard; peptide; 113 AA.
 AC R56139;

DT 29-JUN-1995 (first entry)
 DE Human interleukin-3 mutein (pMon13403).
 KW Human interleukin-3; hIL-3; mutant; mutein.
 OS Synthetic.
 PN WO9412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.

PA (MONS) MONSANTO CO.
 (SEAR) SEARLE & CO G D.
 PI Abrams M, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
 PI Thomas JW;
 DR WPI: 94-200266/24.
 DR N-PSDB: Q68172.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia

PS Claim 32: Page 131; 38pp; English.
 CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
 CC The muteins may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The muteins are made by recombinant
 CC DNA techniques. Recombinant DNA sequences encoding the muteins are
 CC given in GENSEQ Accession Numbers Q68083 - Q68189.
 SO Sequence 113 AA;
 SO 10 A; 5 R; 9 N; 6 D; 0 B; 2 C; 7 Q; 8 E; 0 Z; 2 G; 4 H;
 SO 10 I; 15 L; 4 K; 4 M; 3 F; 10 P; 7 S; 4 T; 1 W; 1 Y; 1 V;

R56139 Length: 113 April 28, 1998 12:48 Type: P Check: 6301 ..

1 MANCISIMIDE IHHHLKRPN PLIDPNLNS EDMILMDSN LRLPNLAFV
 51 HASQOLENAS GIEALIRNLQ PCLPSATAAP SRHPITIKAG DMOEFREKLT
 101 FYLVLEQAO EQQ

11AA_SEQUENCE 1.0
 ID W15514 standard; Protein; 115 AA.
 AC W15514;

DT 28-NOV-1997 (first entry)
 DE Interleukin-3 receptor agonist pMon1155.pcp.

KM Interleukin-3 receptor agonist; stem cell; hematopoietic cell;
 OS human; gene therapy.
 OS Chimeric Homo sapiens;
 OS Chimeric synthetic.
 PN MO9712979-A1.
 PD 10-APR-1997.
 PF 04-OCT-1996; U15941.
 PR 05-OCT-1995; US-004835.
 PA (SEAR) SEARLE & CO G D.
 PI Bauer SC, Baum CM, Caparon MH, Feng Y, McKearn JP;
 PI McWhorter CA, Zurfluhll;
 DR MPI: 97-226222/20.
 DR N-PSDB; T60351.
 PT New peptide(s) with IL-3 receptor agonist activity - used to
 PT stimulate proliferation of haematopoietic cells, or for expansion of
 PT stem cells in vitro
 PS Claim 3; Page 111; 126pp; English.
 CC This polypeptide, designated pMON31155.pep, is a novel interleukin-3
 CC (IL-3) receptor agonist based on human IL-3 (see W14061) and
 CC prepared by multiple PCR of the IL-3 gene and DNA fragment
 CC rearrangement (see also T60351). Claimed IL-3 receptor agonists
 CC (W1514-21) are produced in host cells using a claimed method.
 CC They can be used to stimulate production of haematopoietic cells in
 CC a patient (claimed), e.g. one who has had chemotherapy or
 CC radiotherapy, or is affected by viral infection or some other
 CC immune deficiency. They can also be used for selective ex vivo
 CC expansion of stem cells (claimed) for return to a patient for
 CC treatment of haematopoietic disorders, optionally after insertion
 CC of DNA into the cultured cells. Compared with IL-3, the agonists
 CC may have improved activity (allowing a lower dose) and may also
 CC have reduced side effects and better physical properties such as
 CC solubility, stability and refold efficiency.
 SQ Sequence 115 AA:
 SQ 9 A; 7 R; 9 N; 6 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 5 G; 3 H;
 SQ 10 I; 15 L; 4 K; 2 M; 3 F; 9 P; 7 S; 3 T; 1 W; 1 Y; 4 V;
 W1514 Length: 115 April 28, 1998 12:48 Type: P Check: 8744 ..

1 LDPNNINDE VSIIDNRNLK LPNLESEFVRA VKNLNENAGI EAILRLIQPC
 51 LPSATAPSR HPIIKAGDM QEPREKLFY LVLEQAOQ QGGGSCSIS
 101 IDRIHMLKRP PPAPL

!!AA_SEQUENCE 1.0
 ID R56141 standard; peptide; 106 AA.
 AC R56141;
 DT 29-JUN-1995 (first entry)
 DE Human Interleukin-3 mutein (pMON13412).
 KW Human Interleukin-3; hIL-3; mutant; mutein.
 OS Synthetic.
 PN WO9412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-961044.
 PA (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
 PI Thomas JW;
 DR MPI: 94-200266/24.
 DR N-PSDB; Q68175.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukemia
 PS Claim 32; Page 132; 38pp; English.
 CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-116)hIL-3.
 CC The muteins may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency

CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The muteins are made by recombinant
 CC DNA techniques. Recombinant DNA sequences encoding the muteins are
 CC given in GENESQ Accession Numbers Q68083 - Q68189.
 SQ Sequence 106 AA:
 SQ 9 A; 7 R; 9 N; 5 D; 0 B; 2 C; 3 Q; 6 E; 0 Z; 2 G; 4 H;
 SQ 10 I; 16 L; 4 K; 4 M; 3 F; 9 P; 6 S; 3 T; 1 W; 1 Y; 2 V;
 R56141 Length: 106 April 28, 1998 12:48 Type: P Check: 8813 ..

1 MANCSINIDE ITHILKRPEN PLDPNNINS EDMIDIMERN LTPNLLAFV
 51 RAVKHLENAS GIEAILRLIQ PCLPSATAAP SRHPIIKAG DWQERFLKIQ
 101 FYLSTL

!!AA_SEQUENCE 1.0
 ID R36739 standard; Protein; 245 AA.
 AC R36739;
 DT 13-SEP-1993 (first entry)
 DE Human bone morphogenetic protein BMP-3;
 KW Induction; bone defect treatment; healing; wound; ulcers;
 KW injury; tissue repair; osteoporosis; burns; incisions;
 KW neuronal survival increase; fracture reduction; cartilage growth;
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 142
 FT /note "x represents 229 missing amino acids"
 PN WO9309229-B.
 PD 13-MAY-1993.
 PF 02-NOV-1992; U09430.
 PR 04-NOV-1991; US-787496.
 PR 07-APR-1992; US-864692.
 PA (GEMT) GENETICS INST INC.
 PI Israel D, Wolfinan NM.
 DR MPI: 93-167696/20.
 DR N-PSDB; Q41298.
 PT Recombinant hetero-dimeric BMP proteins - are useful in treating
 PT bone defects, healing bone injury and in wound healing
 PS Disclosure; Page 117; 169pp; English.
 CC The sequence is that of the human bone morphogenetic protein
 CC BMP-3. The 229 missing amino acids represent a page missing from
 CC the sequence as given in the specification.
 SQ Sequence 245 AA:
 SQ 19 A; 17 R; 8 N; 8 D; 0 B; 10 C; 7 Q; 15 E; 0 Z; 18 G; 2 H;
 SQ 10 I; 24 L; 12 K; 5 M; 11 F; 19 P; 22 S; 10 T; 4 W; 8 Y; 15 V;
 1 Others;
 R36739 Length: 245 April 28, 1998 12:48 Type: P Check: 5590 ..

1 MAGASRLTL WLGCRCVSLA QGERPKRPP ELRKAVPGDR TAGGPDSEL
 51 QPODKVSEHM LRLYDRYSTV QAARTPGSLE GGSQPMRPRL LREGTVRSF
 101 RAAAEETLER KGLYFNLTLS LTKSENTLSA TLYFCGEIG NNCARYIKV
 151 DRADIMSEWM IISPKSFAY YCSGACQFPM PKSLAKSNHA TIQSYRAVG
 201 VVFGIEPCPC VPEKMSLSI LFEDENKNV LKYPNMIVE SCACR

!!SEQUENCE_LIST 1.0

(Nucleotide) FASTA of: seq1b from: 1 to: 597 April 28, 1998 08:41

TO: geneseqn:* Sequences: 159,646 Symbols: 57,748,920 Word Size: 6

Sequences too short to analyze: 10 (42 symbols)

Databases searched:

Geneseq-NA, Release 30.0, Released on 26Jan1998, Formatted on 10Mar1998

Searching with both strands of the query.

Scoring matrix: GenRunData:fastadna.cmp

Constant pamfactor used

Gap creation penalty: 16 Gap extension penalty: 4

Histogram Key:

Each histogram symbol represents 814 search set sequences
Each inset symbol represents 4 search set sequences
z-scores computed from opt scores

```

z-score obs exp
(-) (*)
< 20 610 0 : *
22 38 0 : *
24 25 0 : *
26 38 6 : *
28 95 66 : *
30 205 400 : *
32 611 1545 : *
34 1710 4192 : *
36 3933 8609 : *
38 7111 14227 : *
40 11723 19846 : *
42 23395 24259 : *
44 33392 26760 : *
46 48828 27255 : *
48 36046 26094 : *
50 29756 23811 : *
52 25699 20934 : *
54 13874 17881 : *
56 11616 14936 : *
58 9440 12262 : *
60 6740 9933 : *
62 6220 7963 : *
64 4366 6333 : *
66 3549 5006 : *
68 2740 3937 : *
70 2015 3086 : *
72 1633 2411 : *
74 1221 1880 : *
76 839 1463 : *
78 697 1137 : *
80 586 883 : *
82 452 676 : *
84 333 535 : *
86 230 414 : *
88 188 320 : *
90 195 248 : *
92 153 192 : *
94 112 148 : *
96 76 115 : *
98 46 89 : *
100 38 69 : *
102 35 53 : *
104 27 41 : *
106 32 32 : *
108 30 25 : *
110 15 19 : *

```

```

112 18 15 : *
114 2 11 : *
116 0 9 : *
118 8 7 : *
>120 20 5 : *

```

Results sorted and z-values calculated from opt score
1988 scores saved that exceeded 81
125983 optimizations performed
Joining threshold: 52, optimization threshold: 37, opt. width: 16

The best scores are:

```

GENESQN:T49312 Begin: 1 End: 597 Init1 initn opt z-sc E(290156)..
! Human endothelial-monocyte activati... 2985 2985 2985 3056.5 0
GENESQN:086718 Begin: 506 End: 746
! MOUSE EMAP11 CDNA... New endothelia... 424 524 449 452.9 2.2e-18
GENESQN:094113 Begin: 389 Strand: -
! PML2 genomic DNA... Thrombopoietin ... 72 72 136 132.3 1.7
\\End of List

```

seq1b
GENESQN:T49312

```

ID T49312 standard; cDNA; 636 BP.
AC T49312;
DT 04-JUL-1997 (first entry)
DE Human endothelial-monocyte activating polypeptide III encoding cDNA.
KW Endothelial-monocyte activating polypeptide; EMAP III; human;
KW neoplasia; ss. . .

```

```

SCORES Init1: 2985 Initn: 2985 Opt: 2985 z-score: 3056.5 E(): 0
100.0% identity in 597 bp overlap

```

```

seq1b 10 20 30 40 50 60
TACCCTGCTGAAAAAAATGCGCAGCGCTGCTACCCAGATCCCTCAAGCAGAGCC
|||||
T49312 TACCCTGCTGAAAAAAATGCGCAGCGCTGCTACCCAGATCCCTCAAGCAGAGCC
10 20 30 40 50 60

seq1b 70 80 90 100 110 120
AATGCCCAAGGCGCTCCCAAGATTCAAGACAGAGAGAGTATCCATCCCGCGTGAT
|||||
T49312 AATGCCCAAGGCGCTCCCAAGATTCAAGACAGAGAGAGTATCCATCCCGCGTGAT
70 80 90 100 110 120

seq1b 130 140 150 160 170 180
ATCCGTGTGGGAAATCATCACTGTGGAAGACACCCAGATGCACAGCCCTGTATGA
|||||
T49312 ATCCGTGTGGGAAATCATCACTGTGGAAGACACCCAGATGCACAGCCCTGTATGA
130 140 150 160 170 180

seq1b 190 200 210 220 230 240
GAGAGATTGACGTGGGGAAGCTGAACCGAGCTGTGTGAGCGCGCTGTATGTTTC
|||||
T49312 GAGAGATTGACGTGGGGAAGCTGAACCGAGCTGTGTGAGCGCGCTGTATGTTTC
190 200 210 220 230 240

seq1b 250 260 270 280 290 300
GTGCCCAAGGAGAGACTGCAAGCAGCAGCTGTAGTGTGTGTGCAATGAAACCCAG
|||||
T49312 GTGCCCAAGGAGAGACTGCAAGCAGCAGCTGTAGTGTGTGTGCAATGAAACCCAG
250 260 270 280 290 300

seq1b 310 320 330 340 350 360
AAGATGAGAGAGTCCAGTCCCAAGCAGCTTCTGTGTCTTCTAATAGAGGATTAAC
|||||
T49312 AAGATGAGAGAGTCCAGTCCCAAGCAGCTTCTGTGTCTTCTAATAGAGGATTAAC
310 320 330 340 350 360

seq1b 370 380 390 400 410 420
CGCCAGTTGAACTCTGACCCCTCCGCGAGCGCTCTCTCTGTGAGCAGCGTGTGTC
|||||
T49312 CGCCAGTTGAACTCTGACCCCTCCGCGAGCGCTCTCTCTGTGAGCAGCGTGTGTC
370 380 390 400 410 420

```

```

T49312      |||||||
CGCCAGGTTGACCTTGAGCCCTCCGGACGCTGCTCCGTCGAGCAGCTGTTGTG
370          380          390          400          410          420

seq1b      AAGGCTGTGAAAGAGCCACAGATGAGAGCTCAAGCCAGAAAGAGTCTTCAG
430          440          450          460          470          480
|||||
AAGGCTGTGAAAGAGCCACAGATGAGAGCTCAAGCCAGAAAGAGTCTTCAG
430          440          450          460          470          480

T49312      AAGTTCAGGCTGACTTCAAAATTTCTGAGAGTGCATCGCACAGTGGAGCAACACAC
490          500          510          520          530          540
|||||
AAGTTCAGGCTGACTTCAAAATTTCTGAGAGTGCATCGCACAGTGGAGCAACAC
490          500          510          520          530          540

T49312      TTCTATGACCAAGCTGGGCTCCATTTCCGTAAATGCTGAAAGGGGGAACATTAGC
550          560          570          580          590
|||||
TTCTATGACCAAGCTGGGCTCCATTTCCGTAAATGCTGAAAGGGGGAACATTAGC
550          560          570          580          590          600

T49312      CCAGCCGACATCTTCCCCCTTCTTCACCACTGA
610          620          630

seq1b      ID Q86718 standard; cDNA; 1086 BP.
GENESEQ:Q86718
AC Q86718;
DT 29-SEP-1995 (first entry)
DE Mouse EMAP11 cDNA.
KW EMAP11; endothelial monocyte activating polypeptide II; chemotaxis;
inflammation; tissue factor; tumor; cancer; therapy; metha; ...

SCORES      Init1: 424 Initn: 524 Opt: 449 z-score: 452.9 E(): 2.2e-18
65.1% identity in 241 bp overlap

seq1b      GGCCTGCCAAGATTACAGAACCAAGAGAGTATCCATCCCGGCTGGATTCCGTGG
80          90          100          110          120          130
|||||
CGCAGCAGCAGCACTGACTGCTCAAGCCATGACGCGCATCGGCTGTCGAAATG
480          490          500          510          520          530

Q86718      GGAATATCATCTCTGTGAGAAAGCACCCAGATGACAGACGCTGTATGTAGAGAGATTG
140          150          160          170          180          190
|||||
GTTGTATGTTACTCTCCAGAAAGCACCCCTGATGCAATTCATGTATGTGAGGAAGTAG
540          550          560          570          580          590

seq1b      ACCTGGGGGAGACTGACCTGAGCTGTGTGAGGGGCTGTACAGTTCTGCGCCAGG
200          210          220          230          240          250
|||||
ATGTGGGAGAACACAGCCCGCGCAGCGTGTGAGGGGCTGTGATCATGTTCTCTAG
600          610          620          630          640          650

Q86718      AGCAACTGACAGCAGGCTGTAGTGTCTGTGCACTTGAACCCCGAAGATGAGAG
260          270          280          290          300          310
|||||
AACAGATGCAAAATGTATGTGTGTACTGTATCTGTAAAGCTGCAAAAGATGCGGG
660          670          680          690          700          710

seq1b      GAGTGCAGTCCCAAGGAGCTTCTGTGTGCTTCTATTAAGAGATAAACCCCGAGTTG
320          330          340          350          360          370
|||||
GAGTCTGTCTCAAGCAGTGTATGTGTGCTCACTTCAACCAAGAAAGTGAATTTCTG
720          730          740          750          760          770

seq1b      AACCTCTGAGACCCCTCCGAGGCTGTCTGCTCCGTGAGAGCAGTGTGTAAGGCGCTAG
380          390          400          410          420          430
|||||
CCCTCTCCCAACGGGTCCTCTCTGGGACAGATTAATTAATTTGATGCTTTCTCTGAGAC
Q86718

```

```

seq1b /rev
GENESEQ:Q94113
ID Q94113 standard; DNA; 1014 BP.
AC Q94113;
DT 27-FEB-1996 (first entry)
DE pML2 genomic DNA.
KW Human; thrombopoietin; TPO; mpl ligand; hML; fragment polypeptide;
megakaryocytopoietic cytokine receptor; thrombopoietic signal; ...

SCORES      Init1: 72 Initn: 72 Opt: 136 z-score: 132.3 E(): 1.7
60.5% identity in 152 bp overlap

seq1b      CCTCTATAGACGACACAGACAGATGCTGAGGAGCTGAGTCTCTGAGTTCTTGGG
349          359          369          379          389          399
|||||
TGGAGCAGTGTATGACAGACAGGAGGAGCAAGTGGAGCCCTTGCTCTCATTC-CTTCTG
220          230          240          250          260          270

Q94113      TTTGAGT--TGACAG--CACCAGTACCAAGCTGTCTGCAAGTTCTTGG--G
289          299          309          319          329          339
|||||
GTGAGCTTCTGTGACAGGTTGCTCTCTCTCTGCGGCGCCGAGAGCTTGTGATG
280          290          300          310          320          330

seq1b      CACGAACTGATACAGGCGCTGACCAAGTCCGTGTGATCAAGTCTTCCCAAGTAACTT
239          249          259          269          279          289
|||||
CAGGGA--AGGACCA--CAGTCAACAGATCCAGTCCATCTTCTGAACTTCAACAA
340          350          360          370          380          390

Q94113      CTTCATATACAGGCTGTCTGATGTGAGTGTCTTCCACAGTGAATTTTCCACACG
179          189          199          209          219          229
|||||
CTGCTCCGAGGAAGTGTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
390          400          410          420          430          440

seq1b      CPU time used:
Datebase scan: 0:11:13.2
Post-scan processing: 0:00:00.2
Total CPU time: 0:11:13.5
Output File: seq1b.fasta
ID T49312 standard; cDNA; 636 BP.
AC T49312;
DT 04-JUL-1997 (first entry)
DE Human endothelial-monocyte activating polypeptide III encoding cDNA.
KW Endothelial-monocyte activating polypeptide; EMAP III; human;
neoplasia; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT CDS 94..600
FT /tag= a
FT /product= EMAP-III
FT W03640719-A1.
FT P19-DEC-1996.
FT PF 07-JUN-1995; U07328.
FT PR 07-JUN-1995; W0-007328.
FT PA (HUMA-) HUMAN GENOME SCI INC.
FT PI Coleman TA, Olsen HS, Rosen CA;
FT WPI: 97-052219/05.
FT P-PSDB: W06596.
FT New isolated endothelial-monocyte activating polypeptide III - used
PT to develop prods. for the diagnosis and treatment of conditions
PT involving EMAP III e.g. neoplasia
PS Claim 5; Page 37; 50pp; English.
CC The present sequence encodes the 168 amino acids of the novel
CC polypeptide endothelial-monocyte activating polypeptide III (EMAP III).

```

CC The product can be used for the diagnosis and treatment of conditions
CC involving abnormal expression of EMAP III. In particular, EMAP III can
CC be used to regress neoplasia such as in tumours and cancers.
CC EMAP III shows a high degree of homology to EMAP II with 60%
CC identity and 75% similarity over a 150 amino acid stretch.
SO Sequence 636 BP; 172 A; 166 C; 178 G; 120 T;

T49312 Length: 636 April 28, 1998 12:51 Type: N Check: 4493 ..

```
1  TACCCCTGCC CTGAAAAAC TGGCCACGCG TGCTACCCA GATCCCTCAA
51  AGCAGAGGCC AATGGCCAA GGCCTGCCAA GAATTCAGAA CCAGAGGAGG
101  TCATCCCATC CCGGCTGGAT ATCCGTGTGG GGAATATCAT CACTGTGGAG
151  AAGCACCAGC ATGCAGACAG CCTGTATGTA GAGAAATITG ACGTGGGGGA
201  AGCTAACCAC CGAAGCTGGG TGAGCGGCGT GGTACAGTTC GTGCCAAGG
251  AGGAAGTGCA GGACAGGCTG GTAAGTGTGC TGTCAACCTT GAAACCCGAG
301  AAGATGAGAG GAGTCGAGTC CCAAGGCATG CTTCTGTGTG CTTCTATAGA
351  AGGATTAAC CGCCAGGTTG AACCTTGA CCCTCCGACA GGCCTGTGTC
401  CTGTGAGACA CGTGTGTGTG AAGGGCTATG AAAAGGGACA ACCAGATGAG
451  GAGCTCAGAC CCAAGAGAA AGTCTTGAG AAGTTCAGG CTGACTTCAA
501  AATTTCTGAG GAGTGCATCG CACAGTGAA GCMAACCAAC TTCATGACCA
551  AGCTGGGCTC CATTTCTGT AAATCGCTGA AAGGGGGGAA CATGAGTAG
601  CCAGCCAGAC ATCTTCCGCC CTCTTCCAC CACTGA
```

```
!!NA-SEQUENCE 1.0
ID 086718 standard; cDNA; 1086 BP.
AC 086718;
DT 29-SEP-1995 (first entry)
DE Mouse EMAPII cDNA.
KW EMAPII; endothelial monocyte activating polypeptide II; chemotaxis;
KW Inflammation; tissue factor; tumor; cancer; therapy; metha;
KW Sarcoma; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 64..996
FT CD5 /*tag= a
FT PN W0509180-A.
FT PD 06-APR-1995.
FT PR 29-SEP-1994; U11085.
FT PR 29-SEP-1993; US-129456.
FT PA (UNCO ) UNIV COLUMBIA NEW YORK.
FT PI Clauss M, Kao J, Kayton M, Libutti SK, Stern DM;
FT WP: 95-147389/19.
FT P-PSDB: R72577.
FT PT New endothelial monocyte activating polypeptide II - induces
PT chemotaxis, inflammation and tissue factor, useful for treating
PT tumours, also related antibodies, DNA and active fragments
PT Disclosure; Fig.4; 180pp; English.
CC A mouse metha sarcoma cDNA library was screened with a probe
CC based on the N-terminal sequence of mouse EMAPII. Overlapping
CC clones were combined to obtain a contiguous full-length sequence
CC (given in 086718) encoding a 33 kDa protein (R72577).
CC Recombinant EMAPII was expressed in E. coli.
CC Sequence 1086 BP; 346 A; 212 C; 297 G; 231 T;
```

086718 Length: 1086 April 28, 1998 12:51 Type: N Check: 3316 ..

```
1  GAGGCTGCTC AAGAGCTCG GTTGGGTAC CGCTTATGT TTCTGTGCGC
51  ATTCTGGGGA AAGATGGCA CGAATATGC TGTTCGANG AGCGTGAGC
101  AGAAGGTTGC AGAGCGGAT CAGATCATCG AATATCTCAA GCAGCAGGTT
```

```
151  GCTCTTCTTA AGGAGAAAGC AATTTGCAG GCACATGA GAGAAGAAA
201  GAAATTTGGA GTTAAATG CTAACTGAA AAAAAGATA GAAGAGCTAA
251  AGCAAGAGCT GATTCGTGCA GAAATTCANA ACGGAGTGA GCAAGTGCT
301  GTTCATGTA GTACTCCACT GCAGAGAAC TGTACTGCTT CTGAAGTGT
351  GGTGAGTCT CATTCACTAG CAACCAACCG CTCTCTGCTT ACNAAAGAGC
401  AGATCAAGC GGGAGAGAA AAGAGGTGA AAGAGAGAC TGAAGAGAA
451  GGAGAGAAA AGGAGAGCA GCAGTGGCA GCAGCAAGTA CTGACTCAA
501  GCCTATCAGC GCATCGGCTC TGAATCTTG AATTGTTGT ATTTTACTG
551  CCAGAGACA CCTGATGCA GATTCAGTGT ATGTGAGAGA AGTAGATGTG
601  GGAGAGCAG CCCCGGCAC GGTGCTCAG GGGCTGTGTA ATCATGTTCC
651  TCTAAGACAG ATGCAATATC GTATGTGTT TTTACTCTGT AATCTGAAC
701  CTGCAAGAT GCGGGAGAT CTGTCTCAAG CCATGTGAT GTGTGCCAGT
751  TCACAGAGA AAGTGAGAT TGTGGCCCT CCCAAGGGT CCGTTCCTGG
801  GGACGAATT ACTTTGATG CTTTCTCTG AGAGCTGAC AAGAGCTAA
851  ACCCTAAGA GAAGATCTGG GAGCAGATCC AGCCTGACT GCACACCAAT
901  GGTGAGTGG TGCCACATA CAAAGAGCT CCCTTTGAG TGAAGGGAA
951  GGGATTGTC AGAGCCCAA CATTGCCAA TAGTGAATT AAATAGTGC
1001  TCTGTACTG AAGACATTC GCGAAACTT AATAACAATA AAGAGAGTG
1051  TGTATATCAC TTACATATTA AAAAAAAAA AAAAAA
```

```
!!NA-SEQUENCE 1.0
ID 094113 standard; DNA; 1014 BP.
AC 094113;
DT 27-FEB-1996 (first entry)
DE PML2 genomic DNA.
KW Human; thrombopoietin; TPO; mpl ligand; hml; fragment polypeptide;
KW megakaryocytopoietic cytokine receptor; thrombopoietic signal;
KW EPO-domain fragment; erythropoietin; hepo; haematopoietic cell;
KW megakaryocyte; thrombocytopenia; myeloproliferative disease;
KW inflammatory thrombocytosis; iron deficiency; Epo; platelet;
KW red blood cell; progenitor; hml-2; ss.
OS Sus scrofa.
FH Key Location/Qualifiers
FT CDS 1..987
FT CD5 /*tag= a
FT FT /product= PML2
FT PN GB2285446-A.
FT PD 12-JUL-1995.
FT PR 21-DEC-1994; 025831.
FT PR 03-JAN-1994; US-176553.
FT PR 21-JAN-1994; US-185607.
FT PR 15-FEB-1994; US-196689.
FT PR 04-APR-1994; US-223263.
FT PR 25-MAY-1994; US-249376.
FT PR 02-DEC-1994; US-348658.
FT PR 02-DEC-1994; US-348657.
FT PA (GETH ) GENENTECH INC.
FT PI Eaton DL, de Sauvage FJ;
FT WP: 95-234018/31.
FT P-PSDB: R76170.
FT PT Thrombopoietin polypeptide, ligand for mpl cytokine receptor -
PT useful for treating thrombocytopenia and related diseases
PT Example 13; Fig 21; 192pp; English.
CC This sequence represents a genomic clone encoding an isoform of porcine
```

CC thrombopoietin (TPO), also known as mpl ligand (PML), PML2. This
CC sequence was isolated by RACE PCR using the primer sequences given in
CC 094140-43 and 094128. The genomic sequence was isolated from a porcine
CC genomic library in EMBL3 by screening with pR45. Two forms of PML were
CC isolated in this manner. The first was a full length clone (see also
CC 094112) and the second encoded a protein with a four amino acid deletion
CC (PML2). Comparison of the sequences of PML and PML2 shows that the
CC latter form is identical except for the deletion of the tetrapeptide OLPP
CC corresponding to residues 111-114. The four amino acid deletions
CC observed in murine, human and porcine M_L cDNA occur at precisely the
CC same position in the predicted proteins (see also 094110). PML binds to
CC mpl which is a megakaryocytopoietic cytokine receptor which may transduce
CC a thrombopoietic signal.
SQ Sequence 1014 bp; 216 A; 365 C; 220 G; 213 T;

094113 Length: 1014 April 28, 1998 12:51 Type: N Check: 5829 ..

```
1  AGCCCGGCTC CTCCTGCTTG TGACCCCGCA CTCCTAATA AACTGCTTCG
51  TGACTCCCAT GTCCCTCACG GCAGACTGAG CCAGTGCCA GACATTAACC
101 CTTGTGCCAC ACCTGTCTCG CTCCTGCTTG TGGACTTCAC CTGGGGAA
151 TGGAAACC AGACGAGCA GACAAAGCA CAGATGTCC TGGAGGCCAC
201 AACCTTCTG CTGAGGACAG TGATGACAGC ACGGGACAA GTGGAGCCCC
251 CTGCTCTCTC ATCCCTGCTG GTGCACTTT CTGACAGAGT TCGCTCTCTC
301 CTCGGGGCCC TGCAGGACCT CCTTGAATG CAGGGAAGA CCAAGCTCA
351 CAAGGATCCC AGTGCATCT TCTGAACCT CCAACAACGT CTCGAGAGAA
401 AGGTGGCTTT CCTGCTCCTT GTAGTGGGGC CCTCCCTCTG TGCCAAAGAG
451 GCCCCACCG CCATAGCTGT CCCGAGCAGC ACCTCTCCAT TCCACACACT
501 GAACAAAGTC CCAAACAGCA CCTTGATTT GTTGAGACA AACTCCAGTA
551 TCTCAGCCAG AACTACTGGC TCTGATTTT TCAAGAGCT GCAAGCATTC
601 AGAGCCAGA TTCCTGCTCT GCTGAACCA ACCTCCAGGT CCTAGACCA
651 AATCCCTGGA CACAGAGATG GGACACAGG ACCCTTGAGT GGAATTCATG
701 GACTCTTCC TGAACCCCA CCGGGGGCC TCGAGCTCC AGACATTCTT
751 CCAGCAACTT CAGGATGGG CTCCGGGCA ACCTACTCC AGCCTGAGAG
801 GTCTCTTCC CCAGTCACC CTTCTCTTG AGATACACT CTCTCTCTC
851 CTTACCCAC CTTGCCCTCC CCCAGATCC AGCTCCAGC TCTGCTTCTT
901 GACCCCTCTG CGATCACACC CAATCTACG AGTCTCTTC TATTGACAG
951 TCACCTCAT TTCAGAACCC TGTCTCAGA AGAGTAAGT GCTCAGACC
1001 TGCCAACCTC AGCA
```

 WISE (TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1997 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Apr 24 08:28:34 1998; MasPar time 501.50 Seconds

Tabular output not generated. 1407.620 Million cell updates/sec

Title: >US-08-483-534A-1
 Description: (1-597) from US08483534A.seq (2 of 2)

Perfect Score: 597
 N.A. Sequence: 1 TACCCCTGCGCTGAAAAAAC.....TGAAGGGGGGGAACATTAGC 597
 Comp: ATGGGGAGGGGACCTTTTGTG.....ACTTCCCGCCCTTGATACG

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 354530 seqs, 591221332 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database:

eml53
 1:em_in 2:em_or 3:em_com 4:em_ov 5:em_pl 6:em_htg
 7:em_hum1 8:em_hum2 9:em_da 10:em_ro 11:em_un 12:em_v1
 13:em_pat
 genbank105
 14:gb_ro 15:gb_com 16:gb_ov 17:gb_in 18:gb_pl 19:gb_da
 20:gb_st 21:gb_v1 22:gb_ph 23:gb_sy 24:gb_un 25:gb_pat
 26:gb_htg

Statistics: Mean 10.301; Variance 5.091; scale 2.023

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description	Pred. No.
C 1	82	13.7	21948	17	CEP58B3	5.20e-41
C 2	82	13.7	85953	26	CEH04L24	5.20e-41
C 3	82	13.7	131318	26	CEV42E10	5.20e-41
C 4	81	13.6	498	23	ASEMAP11	5.09e-40
C 5	73	12.2	1068	14	MMU01118	4.27e-34
C 6	73	12.2	1086	25	I50092	4.27e-34
C 7	72	12.1	7218	25	I66494	2.46e-33
C 8	66	11.1	1254	14	AF021800	8.32e-29
C 9	34	5.7	1703	18	SCYGL105M	5.20e-06
C 10	34	5.7	2473	24	SCARCI	2.51e-06
C 11	34	5.7	12325	18	SCU11EFT	2.51e-06
C 12	32	5.4	1745	18	SCU11348	4.42e-05
C 13	31	5.2	215	25	I28278	1.80e-04
C 14	30	5.0	11157	19	ECAE000389	7.22e-04

C 15	30	5.0	11000	19	ECOWM67	7.22e-04
C 16	28	4.7	215	25	I28278	1.09e-02
C 17	26	4.4	1663	15	MV92534	1.49e-01
C 18	26	4.4	10558	19	AE001003	1.49e-01
C 19	25	4.2	1131	17	DROHIC02	5.28e-01
C 20	25	4.2	3916	17	DROHGT4P13	5.28e-01
C 21	25	4.2	3991	17	DROHGT3P12	5.28e-01
C 22	25	4.2	4222	17	DROHGT2P11	5.28e-01
C 23	23	4.2	4297	17	DROHGT1PA	5.28e-01
C 24	24	4.0	450	16	CHCKM11	1.82e+00
C 25	24	4.0	4500	17	CEAT0P2	1.82e+00
C 26	24	4.0	10772	17	AE012089	1.82e+00
C 27	23	3.9	565	25	AE04076	6.08e+00
C 28	23	3.9	1000	14	RNRNAMOBB	6.08e+00
C 29	23	3.9	1951	14	MUSRP77B	6.08e+00
C 30	23	3.9	2405	14	RNMOBPR81	6.08e+00
C 31	23	3.9	2486	14	RATMOBPR2	6.08e+00
C 32	23	3.9	3442	14	RNMOBPR1P	6.08e+00
C 33	22	3.7	69	25	I41362	1.96e+01
C 34	22	3.7	722	14	MUSCALMDA	1.96e+01
C 35	22	3.7	1288	19	SPENMA41	1.96e+01
C 36	22	3.7	3113	1	DMR81432	1.96e+01
C 37	22	3.7	3699	17	DMAC000667	1.96e+01
C 38	22	3.7	7869	17	DMKETE1	1.96e+01
C 39	22	3.7	25646	17	PVU51723	1.96e+01
C 40	22	3.7	41088	17	CEH0280	1.96e+01
C 41	22	3.7	82393	17	AC003923	1.96e+01
C 42	22	3.7	98334	26	AC003102	1.96e+01
C 43	22	3.7	128614	26	HS63G5	1.96e+01
C 44	21	3.5	24570	17	CELT05E8	6.10e+01
C 45	21	3.5	186388	26	AC003694	6.10e+01

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	CEP58B3	21948 bp DNA	INV	23-JAN-1998	Rhabdittia; Rhabdittidae; Caenorhabditis elegans	1 (bases 1 to 21948)	Eukaryotes; Metazoa; Nematoda; Secernentea; Rhabdittia; Rhabdittidae; Rhabdittidae; Pelodetidae; Caenorhabditis	Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berts, M., Bonfield, J., Burton, J., Connell, M., Cope, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kersey, J., Kirtlen, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roop, A., Saunders, D., Showkeen, R., Smalton, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wohlman, P.	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans	Nature in press	2 (bases 1 to 21948)	Direct Submission	Submitted (13-MAY-1996) Louis, MO 63110, USA. E-mail: jesusmiller.ac.uk or rvenematode.wustl.edu	Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of clone F58B3. It may be shorter because we only sequence overlapping sections once,		

or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone F5883 is at 1 in this sequence. The true left end of clone ZK809 is at 21836 in this sequence. Coding the sequences below are predicted from computer analysis, using the program Genefinder (P. Green, ms in preparation), and other available information.

The end of this sequence (21836..21948) overlaps with the start of sequence CEZK809.

FEATURES

```

source
1..21948
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="F5883"
complement(join(553..914,967..1094,1525..1892,1939..1989))
/codon_start=1
/product="F58B3.8"
/db_xref="PID:e244447"
/db_xref="PID:g1370035"
/translation="MQLAISFLPFIPIEISICGYLEIHSIOTIOTAVINISNF
PREKSIYHKLKSNVPENYTRIPVFNDRINITIKGKSEDEKERSYKNTPKYL
GLDITIVTVHPRVGLSPAAVARPLUNGIRIYIGCKNMFGEKCEFDENATNSG
KHCNFQKSPCLNGLTGSKCEKLIPLCLCEGVCIDSNPKPICSCSKAYGEKCE
KHKISASLSHYKGFDMFEMFEMOLIFSNLVENELAEIKSEAWETINLLGNVSDT
PFANRNMTETPIITGCVNLDNDIIT"
complement(join(2547..2969,3948..4149))
/notes="protein predicted using Genefinder; Weak similarity
to E.histolytica lysozyme (PIR Acc. No. S55506)"
/codon_start=1
/product="F58B3.1"
/db_xref="PID:e244554"
/db_xref="PID:g1370028"
/translation="MKRFVSALLIGASAARNGIDFIOPVSTATRCQIHKAGSVI
PRFVTSVGLDHTGIONVKARRAAGLTDVGIFFCLRSKCPSAHQVKTDLALSA
GRVSTLMDIERLAMPANASNRATIEAVRRAQVAKQOVGITSYTNMODIVGLDY
HGQSLTMLWATYDGIKDFSKAPFGGWSRPTIHOWEGITTPCGVSDMNYVP"
complement(join(5347..5789,6421..6625))
/notes="Weak similarity to E.histolytica lysozyme (PIR Acc.
No. S55506)"
/codon_start=1
/product="F58B3.2"
/db_xref="PID:e244555"
/db_xref="PID:g1370031"
/translation="MKHFFTLIFCSVGSAAANGIDINGVSTSTFTCIKNGFSPI
IPRIYSSGVDTGVGYONVKNRARAAGLTDVGIFFCLRSKCPSAHQVKTDLALSA
AGKISTLMDIERLAMPADHASNRATIEAVRRAQVAKQOVGITSYTNMODIVGLD
YHGQSLTMLWAPYDGVKDFSKYAPFGGWSKPTIHOWSDTTTPCGVSDMNYIP"
complement(join(8013..8455,8785..8986))
/notes="Weak similarity to E.histolytica lysozyme (PIR Acc.
No. S55506); cDNA EST YK94d3.3 comes from this gene; cDNA
EST YK94d3.5 comes from this gene"
/codon_start=1
/product="F58B3.3"
/db_xref="PID:e244445"
/db_xref="PID:g1370032"
/translation="MKLVSTLILISAVSRNGIDFIOPVSVATFTCVNAGSVFI
PRFVTSVGLDHTGIONVANARAAAGLTDVGIFFCLRSKCPSAHQVKTDLALSA
GRVSTLMDIERLAMPADHASNRATIEAVRRAQVAKQOVGITSYTNMODIVGLDY
HGQSLTMLWAAATDVKDFSKAPFGGWSPTIHOWGTAGCPGCVSDMNYVP"
complement(join(9873..10664,10717..10897,10953..11093,
11447..11539,11655..11759,11810..12324))
/codon_start=1
/product="F58B3.4"
/db_xref="PID:e244556"
/db_xref="PID:g1370033"
/translation="MTLHLETFNFOVIDKRFPAALTDERFSTRKAVDMGRKQKKT
VGNMPLDIELEDEVOPLKTEPKISKTKTDDEDDDEFDEDDDEDLADDEDE
DEISDEENDEEEENKIGMNGFKRLDARGENYDSDSDSDSEDEGDEVEDRDE
GGIELDLANDREVDQERTSNLACNLEMDNKGEDILMLYKSYENRKKTKFKNML
KKEENTFLEIVDFERIEKPOGGSVIGVITVSDGKGLSKEETGPLKLAKAVE
EKREDDMETRTAVREYLVNKLKYYSVTTPGISAVAVYECGDFQFEELGLMD
MRIPDDMDFEEDRVEFLNAEDVNLTKYAKKRSKSAISTGAKILMDEDDPQRKKK

```

CDS

```

FLEAFNGDADGKGLIVDSDESDEANRKTIALMLNTDERSKLDVDMEGEKNGSD
SSDEYVYRDDDEEIGVKKKKKGSDDEDESEVKEKKEDVKTLYGAYKKKOK
LMENKLRKRGTSKEATNTIKTYAADSISKDRDFSLPDSAYALPSSKRRKGLLY
TKOAEKSKSGSSVITETKPEEDLVQKQADQWNRKSKVKN
complement(join(13218..13349,13423..14081,14133..14237,
14568..16021,16068..16391,16444..16523))
/notes="Similarity to yeast methionyl-tRNA synthetase
(SW:SYMC_YEAST): cDNA EST YK9d5.5 comes from this gene;
cDNA EST YK11b4.5 comes from this gene; cDNA EST YK9d5.3
comes from this gene; cDNA EST YK11d4.3 comes from this
gene; cDNA EST YK11d5.5 comes from this gene; cDNA EST
YK190e7.3 comes from this gene; cDNA EST YK190e7.5 comes
from this gene; cDNA EST YK135d9.3 comes from this gene;
cDNA EST YK135d9.5 comes from this gene; cDNA EST
YK472h4.3 comes from this gene; cDNA EST YK472h4.5 comes
from this gene; cDNA EST YK461h12.3 comes from this gene;
cDNA EST YK461h12.5 comes from this gene; cDNA EST
YK354c6.3 comes from this gene; cDNA EST YK354c6.5 comes
from this gene; cDNA EST YK314e7.3 comes from this gene;
cDNA EST YK314e7.5 comes from this gene; cDNA EST
YK287d5.3 comes from this gene; cDNA EST YK287d5.5 comes
from this gene"
/codon_start=1
/product="F58B3.5"
/db_xref="PID:e244446"
/db_xref="PID:g1370034"
/translation="MGHDLADIKKSPFASLPGVEKKDKPSLIPQGRNLTALP
YNNVPHLNTIGCVLSADVPFATYCNLRKHQTFYVGGIDEGTGAETKALQSGCPRE
LCDCKHAIRKIGTYEMFGIDFSHGRTTDHGEICQMDALHKNGYSSQSDOLXC
NOCERKILADRPVYGTCPMCAYDDARDQDQDGGKGLNADVLDAKHMCAKPEVOS
THIFLSDLRLOQKTEHDLRELAKEDNRSSNAVGTAKMKKGLDPRCITRDILKGT
AYLDGEKKVYVWEDAPIGYLSITKCYGNSNAVTKMKNPEVLEFNFVGDVNAFH
ALMFPCSOIGANDNYVNNVLCATELYANEDPKFSKRSRTGIFGDAOCTELPADIR
FILIYRPSQDRTAFSDMPFVYKVNSELNLCNFINRLSVANSFGCVYEDMNTN
DDEYVLESHNECMQMDKOPDVHLDKATKTLNLSRLNQMQATVPVLMKDEEG
KRRAGITIGVANAIVHVSLLYPIPIPTISATIRECGPAPLFTFPICYLKAGKM
ICQPSLEIFLQPAQIAEFKARFGSDOSSAPKTAEPKQKQKAPKDKGKXK
ASIAAEVELEOGAKVISOLIAONLKFKDOAKLFTNQORLDGEKQGLTIVKTLQH
OLIELETTAGISIOKAVPVPVSCPTPTPSAGLITTEAPKKAESPAPSEPKAKKOK
KKGKGAAPVDDTIDVGLDMRVGIIICEKHPDADALYQIDVGEASPRVYSGL
VHVPIDQMNRLVVVLCNLPKPKKGVSSRAMVWCASSPDVYELVNEVADSRPGTVP
VSPPTHRPDEQLNPKKITWETVAEDLKVSAAGFAEWKQPLLIIGSESMTPATLGV
HYK"
join(17716..17921,17974..18388,18443..18919,18977..19057)
/notes="protein predicted using Genefinder; Similarity to
mouse interferon-related protein pC4 (SW:PC4_MOUSE); cDNA
EST cm01c9 comes from this gene"
/codon_start=1
/product="F58B3.6"
/db_xref="PID:e244443"
/db_xref="PID:g1370029"
/translation="MGRNKKKKKDDPYIGPAGKGRNDSDDSSSATYMEDIOSI
MGDVEDMNDVYDQVLDNRADQDKNSTYIDGLHRYNLLTKAPAEFINRKEITMA
VCSRNAKPDSSAEOFLARLVGLVAVOAGEISEMIAPESQKRTILMDSRCVFLTA
CATSLAIVTRICASEDEEVTANAKACRPFAMATKSGSANDGHALTINDVADDPK
ITVLLSNOLEVRILAAETLAFLEHMOARAGRPENKENVLSLMDQVNSSKKTK
KDKKRAQRAVADYDPAFTIADDDADPANNKIQOGLTSLDSCAIKTFYDTCCLLHGL
AQQLLHNEVLRVDFDLGPTALPBAESANKQSLAVHEADKRRNRNVRGQRKRSTVY"
join(19370..19424,19481..19563,20023..20835,20888..21052)
/notes="protein predicted using Genefinder; Similarity to
Arabidopsis DNA-damage-repair protein DRT111
(SW:D111_ARATH)"
/codon_start=1
/product="F58B3.7"
/db_xref="PID:e244444"
/db_xref="PID:g1370030"
/translation="MYNDEDDVPGPPAKQAKPWNHTOMAFMOSQALORAKALQOQ
AAROKLYSSAPPVYDLSRRTKTTTATYTSFQPIRANPSPDINSIFLPKATDESV
MAGEBHVAECETPMPNTYEVAAKEINRQKRETKARVARLDRHEEDKRSKSG
AAIAPPTLMPEPEPPEVFKINTNDEKPPMPSPFPAKGKATSRGLGAAINMK
RGYEGAGLIGSEOGMSATIEKTGVGAGIVAEAPAPFANNSKMAVONAKYIL
QMLNLDLSEVSGEGRKEFADEIKEMEMCCGVVNAVYHVDSEQDEDRQVAVPEFT

```



```

CP      344  GAAACACACAAACACATCCCTGGACTCGACTCTCTCACTCTTCTGGGTTGAGGTTG 285
Db      44246 CAAGACAAACGAAAGCGGTTCTGCATTGATGAGAGTGAAGAGGATGCA 44305
CP      284  CACAGCACACATACACAGCTCTCTGCGAGTTCTCTCTGGGACGACGATGACAGCGCG 225
Db      44306 GAAACTACAGTTGCGGGGCTGACTCTCCAAACATGATCTGTCACGTAAGAGCATCT 44365
CP      224  CTCACACAGTCCGCTGTTGAGCTTCCCGACGATCTCTCTACATACAGGCTGTCT 165
Db      44366 GCATCAGGATCTCTCTCACTTGATGATTGACCAACGCGCATATCAG 44415
CP      164  GCATCTGGGCTCTCTCCACAGATGATTTTCCCGACAGCATATCAG 115

RESULT  4
LOCUS   AEMAPII 498 bp DNA SYN 10-OCT-1997
DEFINITION Artificial sequence DNA for cytokine EMAP.II.
ACCESSION Y08737
NID      91619591
KEYWORDS EMAP.II gene.
SOURCE   synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 498)
AUTHORS Schluesener H.J., Seid, K., Zhao, Y., and Meyermann, R.
TITLE    Localization of endothelial-monocyte-activating polypeptide II
          (EMAP II), a novel proinflammatory cytokine, to lesions of
          experimental autoimmune encephalomyelitis, neuritis and uveitis:
          expression by monocytes and activated microglial cells
          Glia 20 (4), 365-372 (1997)
          97404153
          2 (bases 1 to 498)
          Schluesener H.J.
          Direct Submission
          Submitted (01-OCT-1996) H.J. Schluesener, Institute of Brain
          Research, Calver Str. 3, D-72076 Tuebingen, FRG
FEATURES
  source
    1..498
    /organism="synthetic construct"
    /db_xref="taxon:32630"
    1..498
    /gene="human EMAP.II"
    CDS
    <1..>498
    /gene="human EMAP.II"
    /codon_start=1
    /db_xref="PID:e274145"
    /db_xref="PID:g1619592"
    /translation="SKPIDVSRDLRIGCIITARKHPDADSLVVEVDGEIAPRTV
    SGIVNHYPLEOMONBMVITLCNKRPAKRGVLQAMWCASSPEKIIILAPNGSVPG
    DRITFAFPEBPKEKLNPKKIKMEIQIDPDLHTDECAVITKGVFEVKGAGVCAQTM
    SNGGIR"
BASE COUNT 130 a 111 c 147 g 109 t 1 others
ORIGIN
Query Match 13.6%; Score 81; DB 23; Length 498;
Best Local Similarity 67.5%; Pred. No. 3.09e-40;
Matches 156; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```

```

Db      201  GAAACCGCGGAAATGCGTGCGTCTCTCAGGCAATGCTGATGCGC 251
Oy      291  GAAACCCAGAAATGAGAGGATGAGATGCGATCCCAAGGATGCTGTGTGTC 341

RESULT  5
LOCUS   MMU10118 1068 bp mRNA ROD 14-FEB-1995
DEFINITION Mus musculus endothelial-monocyte activating polypeptide II mRNA,
ACCESSION complete cds.
NID      U0118
KEYWORDS 9498911
SOURCE   house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1068)
AUTHORS Kao, J., Houck, K., Fan, Y., Haehnel, I., Libutti, S. K., Kayton, M. D.,
          Grikscheit, T., Chabot, J., Nowyrod, R., Greenberg, S., Kuang, W.-J.,
          Leung, D. W., Hayward, J. R., Kiesel, W., Heath, M., Brett, J. and
          Stern, D. M.
          Characterization of a novel tumor-derived cytokine.
          Endothelial-monocyte activating polypeptide II
          J. Biol. Chem. 269 (40), 25106-25119 (1994)
          95014290
          2 (bases 1 to 1068)
          Kao, J., Fan, Y., Haehnel, I., Brett, J., Greenberg, S., Clauss, M.,
          Kayton, M., Houck, K., Kiesel, W., Seifeld, R., Burnier, J. and
          Stern, D.
          A peptide derived from the amino terminus of
          endothelial-monocyte-activating polypeptide II modulates
          mononuclear and polymorphonuclear leukocyte functions, defines an
          apparently novel cellular interaction site, and induces an acute
          inflammatory response
          J. Biol. Chem. 269 (13), 9774-9782 (1994)
          94193665
          3 (bases 1 to 1068)
          Kao, J., Ryan, J., Brett, G., Chen, J., Shen, H., Fan, Y., Godman, G.,
          Familletti, P. C., Wang, F., Pan, Y. E., Stern, D. and Clauss, M.
          Endothelial monocyte activating polypeptide II. A novel
          tumor-derived polypeptide that activates host-response mechanisms
          J. Biol. Chem. 267 (28), 20239-20247 (1992)
          93015897
          4 (bases 1 to 1068)
          Houck, K. A.
          Direct Submission
          Submitted (27-MAY-1994) Keith A. Houck, Molecular Biology, Sphinx
          Pharmaceuticals Corp., P.O. Box 52330, Durham, NC 27717, USA
FEATURES
  source
    1..1068
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /cell_line="Weth A"
    /cell_type="fibrosarcoma"
    1..1068
    /standard_name="EMAP II"
    /codon_start=1
    /product="endothelial-monocyte activating polypeptide II"
    /db_xref="PID:g498912"
    /translation="MATNDVAVLRLEBOKGAENDQIEYLRQVALLKERAILOATMRE
    EKLRYENAKLKEIELOELIATIHNGEVRVRLSTNCTGASVVSQSV
    ATASPRATNQIRAGEKRYKTEKEKKEKQQAASSTDKPIDSRLDRIGCI
    VTKKHPDADSLVVEVDGEAPRTVSLVNHVPLEDMQNBMVITLCNKRPAKRG
    VLSQAWMCASSPEKVEILAPNGSVPGDRITFDAPGEPKELNPKKIKMEIQIDP
    DLTNAECVATYKGAPEVKGAGVCAQTMANSIGIR"
    496..993
    /standard_name="EMAP II"
    /product="endothelial-monocyte activating polypeptide II"
BASE COUNT 328 a 212 c 297 g 231 t
ORIGIN
polyA_signal

```


[illegible]

BASE COUNT	577 a	318 c	335 g	473 t	TTNDGLEVFIFKDEEKDHPVVRKLTINAKSESEFKVASIANNQVR"
ORIGIN					
Query Match	5.78;				Score 34; DB 18; Length 1703;
Best Local Similarity	59.58;				Pred. No. 2,51e-06;
Matches 154; Conservative	0; Mismatches 102; Indels 3; Gaps 1;				
Db	1028	CCAGAAAGCCCAAGACCCCAATAGATTTCCGGTTGGTTTATTCGTAAGAACGATC	1087		
Qy	91	CCAGAGGAGGTATCCCATCCCGCGCTGGATATCCGTTGGGCAAAATCATCTGTGGAG	150		
Db	1088	AAGCACCAGATGCGGACTCCCTATATGTCTCTACATATGATCGGTATGAAGAAGC	1147		
Qy	151	AAGCACCAGATGACAGACGCTGTATGATGAAGAAGATGAGGTGGGGAAGCTGAA---	207		
Db	1148	CCAAACTGTTGTTGTGTGGTTGGTCAAGCATTTTTCCTTTGGACCGCTATGCAGAAGCT	1207		
Qy	208	CCAGGACTGTGTGAGGGGCGCTGTACAGTTCGTCGCCCAAGAGAGACTGCGAGGACAG	267		
Db	1208	TATGTGTTGTTGTGATGAACTTGAACAGTTAACATGAGAGGTATTAATCACTGCT	1267		
Qy	268	CTGTGATGAGTGTCTGTGCACTGAAACCCCAAGATGAGGAGTCAGTCCCAAGC	327		
Db	1268	ATGTATTGTGTGTCTTA 1286			
Qy	328	ATGCTTCTGTGTGCTTCTA 346			
RESULT	10				
LOCUS	SCAR1	2473 bp	DNA	UNA	14-OCT-1996
DEFINITION	S.cerevisiae AR1 gene.				
ACCESSION	X95481				
NID	g1620459				
KEYWORDS	AR1 gene.				
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 2473)				
AUTHORS	Simos,G., Segref,A., Fasiolo,F., Hellmuth,K., Shervenko,A., Mann,M. and Hurt,E.C.				
TITLE	The yeast protein Arc1p binds to tRNA and functions as a cofactor				
JOURNAL	EMBO J. 15 (19), 5437-5448 (1996)				
MEDLINE	97050848				
REFERENCE	2 (bases 1 to 2473)				
AUTHORS	Simos,G.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-FEB-1996) G. Simos, University of Heidelberg, Institut fuer Biochemie I, Im Neuenheimer Feld 328, 69120 Heidelberg, FRG				
COMMENT	Related sequence: U31348.				
FEATURES	Location/Qualifiers				
SOURCE	1..2473				
	/organism="unidentified"				
	/db_xref="taxon:32644"				
source	1..2473				
	/organism="Saccharomyces"				
	/strain="J04.2XJR26.198"				
	/db_xref="taxon:4930"				
gene	534..1684				
	/gene="AR1"				
	554..1684				
CDS	/gene="AR1"				
	/codon_start=1				
	/product="Arc1p"				
	/db_xref="PID:g221389"				
	/db_xref="PID:g1620460"				
	/translation="MSDIYMKPESLIISKPYSTKQSAQAAQWESYVLSGQIQPHI				
	DQNLVLRNDTFIVSTIPIPTSDVHFEVALPLIKDLVASSKDVSTYTYRHLRLWD				
	DIYQNLLEVSSTDKLEINHDLDPHEVIERKKRPAAGAADAAKADSDVSKKAKKOD				
	HPGKDEDFLTKILREBAKAKAARAKAANKAQQOQENKAPKPKPSAIDPFRVGIQR				
	AIHNPADSLIVSTIDVDEEGEPYVCSGLVKEHPIDAMQERYVVVVCNKIPVNRGI				

BASE COUNT	837 a	457 c	494 g	685 t
ORIGIN	KSTAMVLCGSSDDKVEEVEPPPKDSKADNYFFEEFGDAPPKQINPKKKTMEHQLPHF TTNDGLEIVIFKDEEEKDHPVKRLKLNAGGSESEKVASIANAQYR"			

Query Match	Best Local Similarity	5. 7%;	Score 34;	DB 18;	Length 12325;	
Matches	154;	Conservative	0;	Mismatches 102;	Indels 3;	Gaps 1;
Db 10634	CCAGAAAGCCAAAGCCATCCGCAATAGATTTCGCTGTGGTGTATTATTCACAAAGCCATC	10693				
0Y	91 CCAGAGAGAGTATCCATCCATCCCGCGCTGGATATCCGTGGGGAATAATCATCATCTGGAG	150				
Db 10694	AAGCACCAGATGCCGACTCCCTAATATGCTCTACAAATTGATGTGGGTGATGAGAGAGGC	10753				
0Y	151 AAGCACCAGATGACAGACGCTGTATGTAGAGAAAGATTGACGTGGGGGAACCTGAA---	207				
Db 10754	CCAAACACTGTTTGTCTTGCTTGGCTCAAGCAATTTTCCTTGGAGCGATATGCAAGAAAGT	10813				
0Y	208 CCAGGACTGTGTGATGAGCGGCTGGTACAGTTCGTGTCGCCAAGAGAGAACTGACAGACAG	267				

[illegible]

Oy 328 ATGCTCTGTGCTTCTA 346
 RESULT 13 128278 215 bp DNA PAT 30-OCT-1996
 DEFINITION Sequence 5 from patent US 5569830.
 ACCESSION 128278
 MID g1819054
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE
 1 (bases 1 to 215)
 Benetti, A., Labavitch, J.M., Powell, A. and Storz, H.
 TITLE Plant inhibitors of fungal polygalacturonases and their use to
 JOURNAL control fungal disease
 FEATURES Patent: US 5569830-A 5 29-OCT-1996;
 source Location/Qualifiers
 1..215
 BASE COUNT 15 a 8 c 25 g 26 t 141 others
 ORIGIN
 Query Match 5.2%; Score 31; DB 25; Length 215;
 Best Local Similarity 14.3%; Pred. No. 1.80e-04;
 Matches 22; Conservative 64; Mismatches 67; Indels 1; Gaps 1;
 Db 1 MTNVMSSSSVYRSASCDKAKKNDGNTSSWTDDCNRTMGVCDDTDTYRNNDGHNK 60
 Oy 395 GTGCTCCGTGTGACACGCTGTTGTGAAGGCTATGAAGAAGGCCAACCGATGAGGACC 454
 Db 61 YSSANYNGNNGVGAAKTHYHTNVSAGSKVTDSYNASTSSNGTSDGNRSAGDSY 120
 Oy 455 TCAACCCCAAGAAGAAGAGCTTCGAGAAGCTTCAGAGCTACTCAAAATTTCTGAGAGAT 514
 Db 121 GSSKTAMTSRNFTGKTANNVDSRNMGASVGS 154
 Oy 515 GCATCGCACAC-GTGGAGCAACCAACTTCATGA 547
 RESULT 14
 LOCUS ECAGE000389 11157 bp DNA BCT 02-SEP-1997
 DEFINITION Escherichia coli K-12 MG1655 section 279 of 400 of the complete
 genome.
 ACCESSION AE000389 U00096
 MID g1789451
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Escherichia coli.
 Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE
 1 (bases 1 to 11157)
 Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, Y.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J.J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mau, B. and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 REFERENCE
 2 (bases 1 to 11157)
 Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 3 (bases 1 to 11157)
 Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

COMMENT

The E. coli K-12 sequence and its annotations have been updated.
 All of the ambiguous residues in our original submission have been
 resolved, and mis-assemblies in two repetitive regions have been
 realigned. The annotations have been improved and updated as well.
 With this release we begin designating a version number for the
 annotated sequence, to assist in keeping track of corrections,
 updates, and other changes. This is version M52 (SEPT. 02, 1997).
 In addition, a revised notation has been instituted which assigns
 each gene (protein- or RNA-encoding) a unique numeric identifier
 beginning with a lowercase 'b' (in the 'label' field); this will
 remain constant through further updates, gene identifications, etc.
 This sequence was determined by the E. coli genome project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHRR). The entire sequence was independently
 determined from E. coli K-12 strain MG1655.
 Predicted open reading frames were determined using Genemark
 software, kindly supplied by Mark Borodovsky, Georgia Institute of
 Technology, Atlanta, GA, 30332.
 e-mail: mark@amber.gatech.edu
 Open reading frames that have been correlated with genetic loci are
 being annotated with CG Site NOS., unique ID nos. for the genes in
 the E. coli Genetic Stock Center (CGSC) database at Yale
 University, kindly supplied by Mary Berlyn. A public version of the
 database is accessible (<http://cgsc.biology.yale.edu>).
 Annotation of the genome is an ongoing task whose goal is to make
 the genome sequence more useful by correlating it with other data.
 Comments to the authors are appreciated. Updated information will
 be available at the E. coli Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>).

FEATURES

source

1..11157
Location/Qualifiers

/organism="Escherichia coli"

/strain="K-12"

/sub_strain="MG1655"

/db_xref="taxon:562"

174..797

/note="0207: This 207 aa ORF is 42 pct identical (7 gaps)
to 52 residues of an approx. 200 aa protein SLTD_ECOLI SW:
P30856"

/codon_start=1

/label="b3071"

/db_xref="pid:g1789452"

/transl_table=1

/translation="MSHHBCCCHHEGQPRHEGCKEKEKSHHEGCHGHEHGGCCG
GRHGGGGRORFEGHGLRVITLIDILSRDSDSHYELIKAIENLTGQNPSPVITP
TLDFLEQOSLITREEGGRKQIALTEGQAGMLEENREOVMIEIRIKARCVGALRO
NPQMRALDNFKAVLDLRVNSDISDAQIKIIVIDRAAFDITQLD"

complement(951..2471)

/gene="air"

complement(951..2471)

/gene="air"

/note="f506: 100 pct identical to AIR_ECOLI SW: P50466"

/codon_start=1

/label="b3072"

/product="aerotaxis receptor protein"

/db_xref="pid:g1789453"

/transl_table=1

/translation="MSHPPVTOQNTPLADDTLTMTDLSQYITTHANDTFVQSGYT
LOELQGPNNVVRHDPMPKAFADMPYTLKGEPSGIVKRRKNGHGVRAVAVM
VREGKISGYSIFRTADEIAAEPIYKALNAGRTSKRIHGLVYRGWGRKPSLP
LRRARVMTLMTLLAAMLFPVAPVYITCALVYLLASACEPMQIVPIEVAIO
ALKVATGERNSVEHLNDSDELGLTRAVGOLGICRWLINDVSGVSSVANGSTELAK
GTDELNEHTOOTVDNVQOYVATNMQASVKONSATASADKLSITASNAVGGELM
TTVIKTDNDIASTORIGTITSLINDIAFTNTIALNAVEAARGOGGFAVAGE
VRHLASANAANDIRKLIDASADKVOGSQOYVAAQRTMEDIVYAVKNTQILAQIS
HSTLEQADGLSLFRAYDELNLITQKNAELVEEQAQSAVVKHRSLEDAVYTLH"

2778..4268

/gene="y9j6"

2778..4268

/gene="y9j6"

/note="2.6.1.13"

/EC_number="2.6.1.13"

/note="0496: 100 pct identical to 429 amino acids of


```

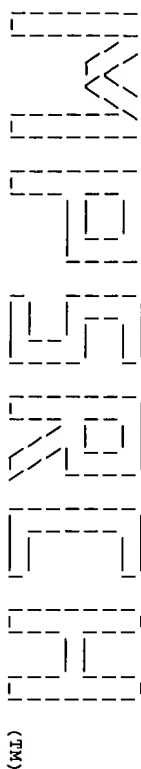
/sub_strain="MG1655"
/notes="This sequence comprises the following lambda
clones: D6693(EC30K511-4), D6695(EC27-779),
D6698(EC30K513A-10pp), D6701(EC30K514A-6pp),
D6703(EC30K514A-4pp), D6704(EC27-355), D6707(EC30K518-4),
D6709(EC30K518-5), D6712(EC17-155), D6713(EC30K520-13pp),
D6715(EC30K520-15pp), D6718(EC27-977), D6721(EC30K523-9),
D6723(EC30K524-3), D6725(EC27-1316), D6727(EC30MM23),
D6730(EC27-417), D6731(EC18-29), D6734(EC19-161),
D6737(EC5-106), D6741(EC30TW4), D6743(EC27-1299),
D6744(EC29-14), D6747(EC22-12), D6750(EC17-214),
D6751(EC18-283), D6753(EC17-149), D6757(EC14-35),
D6758(EC17-5), D6760(EC27-516), D6762(EC30K619-15pp),
D6763(EC30MM35), D6765(EC30MM31), D6766(EC27-394), and
D6770(EC18-109); M3mp19 or Janus vectors were used for
subcloning"
/db_xref="taxon:562"
/map="67.4 to 76.0 minutes"
/notes="predicted bend of 75.70 degrees"
1. .112
/notes="ORF_c496"
136. .1626
/codon_start=1
/db_xref="PID:9606011"
/transl_table=11
/transl_table=11
/translation="MITEFVPIPIPAAGVAOSLOYLNRHYVIREPPEHILNRLPS
ASALACSAHALNLEKRLDHEEMKALREVIEYERHPGFELEKRSYAGDGYA
VEMQAGSINTLADPGOGFIDLCGFGFENNGHRPVYSAQONLQAPLHSDILD
PLRALAKTLALDTPGKLKISFSGSIESVEALKLAKATOSPGRKTFATSGAFH
GKSGALSAATKSTFRKPFMPLPGFRVPGFENIEMRTALNECKRTGDVAAVILC
IOGEGVILPPGYLTAVRKLDDEFGALMIDVOTGMRGKACHEWVQDILC
LAKALGGVMPGIIATIEEVSVLFDPNPFHTTFFGNGPLACAALATINLEONL
PVAOEOKGMDLDFGROLAREYDPLVOEARGGMALIEFYDNEIGYAFASMEFRQY
LVAQNLNAKTRILREPLTLTIECELYIKAKRKLAAARVSEEA"
complement(1668. .2000)
/notes="ORF_f110"
/codon_start=1
/db_xref="PID:9606012"
/transl_table=11
/translation="METVAVADPARLEMRVGKIVERRHENDKLYIYOVDGOKTIO
TTSIIVPYSEELMKGTVVVLNQLAKAKMGISECMILCAETDDGSESVLTPERA
MPAGRVV"
2094. .7072
/notes="corresponds to GenBank Accession Number M6441,
ECOEBCGRA"
2094. .7072
/notes="corresponds to GenBank Accession Number X52031,
ECBE6"
2219. .6461
/notes="corresponds to GenBank Accession Number X03228,
ECBEBCRA"
2219. .3202
/notes="ebgr"
2219. .3202
/notes="ebgr"
/notes="CG Site No. 829"
/codon_start=1
/function="ebg repressor"
/db_xref="PID:9606013"
/transl_table=11
/transl_table=11
/translation="MALKDIAIEAGVSLAVSRVLDNPTLVNKEETKRRHILLETAK
LEYTSARKLOTGAVNOHILAIYSOOEELNDPYLAIRHGJETOCERKGIELN
CYEHSGLDINKNGILIVGPTPALRAASALNDICFIDHPERGSGCYADVILAR
ISKELIDFYNIGVNRIGFEGEDGPRGADIRAVAFAYKRLKOVYRREDIWRGFSF
SSGVELAKOMLAREDPKALFYASDSTIGVRAIHERGNIPODISLISVNDIPTAR
FTFPLPSTVRIHSEMGSGVNLVYERAKDRAPLVLVPSKILKRGTR"
3231. .3383
/notes="predicted bend of 87.49 degrees"
3350. .6478
/notes="ebga"
3350. .6478
/notes="ebga"
/notes="CG Site No. 830"

```

```

/codon_start=1
/product="phospho-beta-D-galactosidase; alpha-subunit"
/db_xref="PID:9606014"
/transl_table=11
/translation="MNSDSLSPGCPNKNENIOLTHENRILAPPAIFPSIDVAQART
FARTSLFLPLSGQNMHFFDHPLOYEATSLMDMGIITVPAMQMGHKLQY
TDEGFPPIADPEFSDNPTGAYORIFTLISGMQKOTLIFDVEVYFEVYNGOVY
GFSGSPRIATFEDISAWKTDNLCYRVMQADSTVEODMMMSAGIFDYLYGK
HLTHINFTVDEDEAYCOATLCEVYENLAASPVTLTEYLTPGERGERSAID
HLAIEKLSAFATVPOQMSAESYLYHLVTLNDANGNVLEVYQARGFEDIKY
RDGLFWNNRTVNLGVNRHNDNRKGAQMDVEVDLQMLKOHNTNSVTAHPND
PRFELDIYOLFVMASTDVESHQFANVDISRTDPOWKEVVERIVRIHAQKNH
PSIILWSDGNSGYGCNIRAMVHAAKALDRLVYHEDRDAEVVDIISTWYVPLM
NEFEYPPKRIICERAHAMNGPGLTEYONVYKDCIOGHVWMCDSGIOADQ
DHGVWVKFGSDYGDYNNVNFCDGLIYSOPGPGRLKEVQYLAIPKIHARDLGB
ELVKNIMFTTLDITLHAVERAEGELTAOQKLDVAPNSAPLOITIPOLARE
AFLNITVTKDSRTYSEAGHPITAYQPLENTQAPVPFVANNRPLLEDRLSCYV
RGYNFAITFSKSGKPTSMOYNGESLTLREKINFPKMDNHRKOEYEGLMQPHLQI
MOEHLRDFAVEQSDGEVLLISRTVIAPEVPEFGRCYIMRIADGQVNVALSERYG
DYPIITICIGFTMGINGEYDQVAYVGRGPGENYVDSOQANTIDIRQAVDMPENYD
PQNNGNRQHWRTALTRHNGGLVYQRPINFSAMHYTOENITHAACHNLOSDOI
TLNLDHQLGLGDSMSGEVLDNWRVWFDFSTGFTLLPVSGEATQSLASTFEFGAG
FESTLHSENKQ"
6475. .6924
/notes="ebgc"
6475. .6924
/notes="ebgc"
/notes="CG Site No. 18412; difference near end makes this
25 amino acids shorter than previously published
sequences"
/codon_start=1
/product="phospho-beta-D-galactosidase, beta-subunit"
/db_xref="PID:9606015"
/transl_table=11
/translation="MRLIINDLEQROIYASGKKQRCVEALENTDNIQPVASHIGDS
LTYREVDSADALFTGHRRYFEVHYVLOGQKLEYAPKETLQVVEYRDETDEYIK
GCGETVEVHEGOIYCIDHEAYRFICNNAVKVLKVTIEDGYFHNK"
6987. .8420
/notes="ORF_o477"
/codon_start=1
/db_xref="PID:9606016"
/transl_table=11
/translation="MSDKRNTIGKFGLLSTFEAAVYSFNNVINNNIELGLASMPF
LATIFVPIPLCLIAEFVSLKNSEAGVAVLNVKSLGRNAFIRAYVWFNLEFES
LLPRVIAVASYAFEGYIMTPVATLISMLTAFSTVWNGKMGKITSVSTLM
LLTLSTILLAGTLVGGVQPADITTDAMIPNFMPLGTTTIRPAAGAEVAY
VNDVKGSKSVKVIILAGLFTGVLYSVSLVNLVYSSKELFTGSGVGHAMAY
FGLPEALMNRVGLVSTFAMFGLMTATPVKLFSEIPGIGKKTVELNENGVA
RRSYDVPDHRPADYVPLGSMNVQDMMNTIINNTAASMLPLFIMLAYINLRKLD
HLPDFNRGSRRTGIIVYSLIAIFAVGFVASTPFGANILITIFYVNGVIYVLFGR
MKVSKYIKGTLAEERHIEATPASNV"
8554. .9624
/notes="ORF_o356"
/codon_start=1
/db_xref="PID:9606017"
/transl_table=11
/translation="MKLTAPCRALLAPPCYARSAGEARPAEHDDTPATSTS
SPSFRFYGELGVGGMIDLEGNKRYKIDGTIIEGLMKTSFGLIYBQWYQAOH
DGNAMVPDHSWGGFEGGINRFGYGRNDGELISLRQSSLDLQWMDGFTPDLYQ
VIPNTRDIMEITAKYONLSGNFRYSVATAPAGHNBESKAMLFGRYDIDKRYVPAH
NGYLOYDLAEITWNLGELITDGGOLYLGILLPNAARWHTTGADGADGVGSS
GMVYSAVNEALIKGYLLSTAYTAARHDDDETTSMOGIWEYGGGRATVAFDSR
PYKMAHSDPDSQFLMOYTV"
9641. .11992
/notes="ORF_o783"
/codon_start=1
/db_xref="PID:9606018"
/transl_table=11
/translation="MKIKTILTPYTCALLSFSAAHANADNYKAVNRTGAPQYMKY
DYDDHQRPNPFDDGAWGHLLPDGNTMGFGVALLTIEYINFTASNDFRLTWOD
GKRVDTLEASIPGALVOKUTADVOVENTLARPRTSLLETKITSNPFLLDWOG
ELIEKLEAKEGKPLSDKTIAGEYDQYKISATSDGLKGVFGKVRATWDLITGSESY
QVHKSILPVQTEINGNRTSKAHINGSTLYTYSHLLTAAEVSEKQAIIDILARPAT

```

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch.un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Apr 24 08:20:39 1998; Maspar time 349.28 Seconds
1215.222 Million cell updates/sec

Tabular output not generated.

Title: >US-08-483-534A-1
Description: (94-597) from US08483534A.seq (1 of 2)
Perfect Score: 504
N.A. Sequence: 94 GAGGAGGTCATCCATCCCG.....TGAAAGGGGGAGACATTAGC 597
Comp: CTCCTCCAGTAGGAGGAGGC.....ACTTCCCCCTTGTAATGC

Scoring table: TABLE default
Gap 6

Nmatch STR: Dbase 0; Query 0

Searched: 1126798 seqs, 421087984 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-est
1:em_est1 2:em_est2 3:em_est3 4:em_est4 5:em_est5
6:em_est6 7:em_est7 8:em_est8

Database: genbank-est
9:gb_est1 10:gb_est2 11:gb_est3 12:gb_est4 13:gb_est5
14:gb_est6 15:gb_est7 16:gb_est8 17:gb_est9 18:gb_est10
19:gb_est11 20:gb_est12 21:gb_est13

Statistics: Mean 10.346; Variance 1.999; scale 5.176

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	438	86.9	475	19	AA152266	2104c02.r1 Soares preg 0.00e+00
2	402	79.8	533	11	H10555	ym04e04.r1 Homo sapien 0.00e+00
3	289	57.3	407	10	RI8984	yg25d10.r1 Homo sapien 0.00e+00
4	278	55.2	459	15	AA102052	2178g12.r1 Stratagene 0.00e+00
5	275	54.6	525	15	AA086129	2184e05.r1 Stratagene 0.00e+00
6	267	53.0	337	17	AA154293	mq49e07.r1 Soares 2NM 0.00e+00
7	219	43.5	243	14	N87033	LI914F fetal heart, La 0.00e+00
8	215	42.7	423	14	AA003398	mg57b09.r1 Soares mous 0.00e+00
9	202	40.1	517	17	AA221081	mv65f03.r1 Soares mous 0.00e+00
10	193	38.3	248	18	AA327316	EST30628 Colon I Homo 0.00e+00
11	163	32.3	209	11	H22113	Y134g06.r1 Homo sapien 7.75e-276
12	157	31.2	500	14	AA049626	mj36d06.r1 Soares mous 3.33e-263
13	157	31.2	510	14	AA051121	mg74c08.r1 Soares mous 3.33e-263
14	113	22.4	437	11	H13315	Y172b07.r1 Homo sapien 8.64e-172
15	104	20.6	470	15	AA102053	2178g12.s1 Stratagene 1.89e-153

16	100	19.8	308	19	AA369979	EST81510 Prostate gland 2.37e-145
17	100	19.8	503	17	AA230715	mw04g08.r1 Soares mous 2.37e-145
18	78	15.5	362	23	R1CC02054A	Rice cDNA, partial seq 1.49e-101
19	73	14.5	587	15	AA120159	mn33c11.r1 Beddington 8.03e-92
20	73	14.5	691	17	AA218315	mq75a01.r1 Soares mous 8.03e-92
21	69	13.7	578	17	AA138243	vb13a01.r1 Stratagene 4.12e-84
22	67	13.3	522	18	AA289897	vb13a01.r1 Soares mous 2.76e-80
23	65	12.9	438	16	AA198745	mv39a01.r1 Soares mous 1.77e-76
24	64	12.7	506	14	W30441	mc21c12.r1 Soares mous 1.39e-74
25	64	12.7	508	17	AA242568	mx24e07.r1 Soares mous 1.39e-74
26	56	11.1	338	22	AA435359	ve13d08.r1 Soares mous 1.33e-59
27	53	10.5	193	12	N50217	yy79f05.r1 Homo sapien 3.83e-54
28	53	10.5	441	17	AA268851	va44c02.r1 Soares mous 3.83e-54
29	50	9.9	619	15	W90821	mf78d07.r1 Soares mous 1.01e-48
30	46	9.1	360	22	C13934	C.elegans cDNA clone Y 1.26e-41
31	46	9.1	411	14	AA033217	mi37d12.r1 Soares mous 1.26e-41
32	44	8.7	420	14	W75273	me33d03.r1 Soares mous 3.88e-38
33	44	8.7	480	14	W75295	me51g01.r1 Soares mous 3.88e-38
34	44	8.7	495	14	W75249	me53c04.r1 Soares mous 3.88e-38
35	43	8.5	427	23	AA456374	aa14e06.s1 Soares NHM 2.06e-36
36	43	8.5	454	23	AA456533	aa15b05.r1 Soares NHM 2.06e-36
37	37	7.3	238	16	AA173393	zp47c10.r1 Stratagene 2.48e-26
38	37	7.3	428	15	AA134849	zn90e06.r1 Stratagene 2.48e-26
39	35	6.9	475	23	AA455165	aa15b05.s1 Soares NHM 4.28e-23
40	35	6.9	487	23	AA455908	aa14e06.s1 Soares NHM 4.28e-23
41	33	6.5	295	22	C03071	Human Heart cDNA, clone 6.23e-20
42	33	6.5	317	22	C03104	Human Heart cDNA, clone 6.23e-20
43	29	5.8	183	15	AA090646	Y1128.seq.F Fetal heart 7.31e-14
44	28	5.6	428	19	AA150204	Z104c02.s1 Soares preg 2.09e-12
45	27	5.4	141	16	AA064500	ml50e08.r1 Stratagene 5.57e-11

ALIGNMENTS

RESULT 1
LOCUS AA152266 475 bp mRNA 14-MAY-1997
DEFINITION Z104c02.r1 Soares pregnant uterus NBHPV Homo sapiens cDNA clone
491330 5' similar to TR:G498912 G498912 ENDORHELIAL-MONOCYTE
ACTIVATING POLYPEPTIDE II. ;

ACCESSION AA152266
NID g1721603
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Euthera; Primates; Catarrhini; Homidae;
Homo.

REFERENCE 1 (bases 1 to 475)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevisan,E.,
Waterson,R., Williamson,A., Wohlmann,P. and Wilson,R.

Washu-Merck EST project
Unpublished (1995)

TITLE

JOURNAL

COMMENT

Contact: Wilson RK
Washu-Merck EST project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNC; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 1171 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 431.

FEATURES

SOURCE

1..475
/organism="Homo sapiens"
/note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'

rsite2-Hind III whole brain from a 73 days post natal female. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'] AACTGAGAAATTCGCGCCGAGGAAATTTTCTTTTCTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lnfmd BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Patima Bonaldi.

ORGANISM
Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 407)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE
The Washu-Merc EST Project

JOURNAL
Unpublished (1995)

COMMENT
GDB: G00-406-036
Contact: Wilson RK
Washu-Merc EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 294
Source: IMAGE Consortium, LML
This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Location/Qualifiers
1. 407
/organism="Homo sapiens"
/clone="33689"

BASE COUNT
103 a 87 c 121 g 95 t 1 others

ORIGIN

Query Match 57.3%; Score 289; DB 10; Length 407;
Best Local Similarity 94.5%; Pred. No. 0.00e+00;
Matches 364; Conservative 0; Mismatches 10; Indels 11; Gaps 11;

Db 1 AGATTGACGTGGGGAGAGTGAACCAAGAGTGTGTGAGCGGCTGTGACAGTTCGTG 60
|||
Oy 184 AAGATTGACGTGGGGAGAGTGAACCAAGAGTGTGTGAGCGGCTGTGACAGTTCGTG 243
|||
Db 61 CCCAAGAGAGACTGCGACGACAGGCTGTGTGCTGTGCAACTGAAACCCCAAG 120
|||
Oy 244 CCCAAGAGAGACTGCGACGACAGGCTGTGTGCTGTGCAACTGAAACCCCAAG 303
|||
Db 121 ATGAGAGAGTGGAGTCCCAAGGATGCTGTGTCTCTATAGAAAGGATAAACCCG 180
|||
Oy 304 ATGAGAGAGTGGAGTCCCAAGGATGCTGTGTCTCTATAGAAAGGATAAACCCG 363
|||
Db 181 CAGTTGAACTTGTGACCTCCGCGAGGCTTGTCTCTGCGTGAACAGTTCGTG 240
|||
Oy 364 CAGTTGAACTTGTGACCTCCGCGAGGCTTGTCTCTGCGTGAACAGTTCGTG 420
|||
Db 241 AAGGCTATGAAAAAGGCCCAACGATGAGAGCTCAAGCCCAAGAAAGTCTTGA 300
|||
Oy 421 AAGGCTATGAAAAAGGCCCAACGATGAGAGCTCAAGCCCAAGAAAGTCTTGA 479
|||
Db 301 GAAGTTGCGAGGCTGTGCTTTCAAAATTTCTTGAGAGTGCATCGCAAGTGGGAAGCA 360
|||
Oy 480 GAAGTTGCGAGGCTGTGCTTTCAAAATTTCTTGAGAGTGCATCGCAAGTGGGAAGCA 533
|||
Db 361 AACCACTTTCATGACCAAGTGGG 385
|||
Oy 534 AACCACTTTCATGACCAAGTGGG 557
|||

RESULT 4

LOCUS AA102052 459 bp mRNA EST 28-OCT-1996
DEFINITION 2178412.r1 Strataene colon (#937204) Homo sapiens cDNA clone 510790 5' similar to WP:F58B3.5 CE06007 METHIONYL-TRNA SYNTHETASE

ACCESSION AA102052
NID 91645892
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eucaryotae; mitochondria eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 459)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE
Washu-Merc EST Project

JOURNAL
Unpublished (1995)

COMMENT
Contact: Wilson RK
Washu-Merc EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28W13 rev2 from Amersham
High quality sequence stop: 418.
Location/Qualifiers
1. 459
/organism="Homo sapiens"
/note="Organ: colon; Vector: pBluescript SK-; Site:1: EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo dT. T-84 colonic epithelial cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."
/db_xref="taxon:9606"
/clone="510790"
/lab_host="SOLR cells (kanamycin resistant)"
<1..>459

BASE COUNT
114 a 118 c 117 g 107 t 3 others

ORIGIN

Query Match 55.2%; Score 278; DB 15; Length 459;
Best Local Similarity 95.6%; Pred. No. 0.00e+00;
Matches 368; Conservative 0; Mismatches 2; Indels 15; Gaps 15;

Db 1 AGGGGCTGTGTACAGTTCGTGCGCAAGAGAGAACTGACAGACAGCTGTGTGTCGTG 60
|||
Oy 223 AGGGGCTGTGTACAGTTCGTGCGCAAGAGAGAACTGACAGACAGCTGTGTGTCGTG 282
|||
Db 61 TGCACTGAAACCCCAAGAGATGAGAGAGTGGAGTCCCAAGCAATGCTGTGTGTCGTG 120
|||
Oy 283 TGCACTGAAACCCCAAGAGATGAGAGAGTGGAGTCCCAAGCAATGCTGTGTGTCGTG 342
|||
Db 121 TCTATGAAAGGATTAACCCAG-TTGAACCTGTGACCTCCGCGAG-CTCTGCTCT 178
|||
Oy 343 TCTATGAAAGGATTAACCCAG-TTGAACCTGTGACCTCCGCGAG-CTCTGCTCT 402
|||
Db 179 GGTGAGCAGCGTGTGTGTAAGGCTATGAAAGGCCCAACAGATGAGAGCTCAAGCC 238
|||
Oy 403 GGTGAGCAGCGTGTGTGTAAGGCTATGAAAGGCCCAACAGATGAGAGCTCAAGCC 461
|||
Db 239 CAGGAAGAAGCTTTGAGAAAGTTCAGAGCTGACTCAAAATTTCTTGAAAGAGTGCAT 298
|||
Oy 462 CAGGAAGAAGCTTTGAGAAAGTTCAGAGCTGACTCAAAATTTCTTGAAAGAGTGCAT 518
|||
Db 299 CACCAAGTTGGAACCAACCTTTCATTTGACCAAGCTGGGCTCATTTCCTGTTA 358
|||

```

519 CGACA-GT-GGAAGCAAAACCAACTT-CAT-GACCAAGCT-GGGCTCATTTC-TGT-A 571
Db 359 AATCGCTGAAAAGGGGGGAACATT 383
||||| ||||| ||||| |||||
QY 572 AATCGCT-GAAA-GGGGGGAACATT 594

RESULT 5
LOCUS AA086129 525 bp mRNA EST 23-OCT-1996
DEFINITION Z184e05.r1 Strataene colon (#937204) Homo sapiens cDNA clone
511328 5' similar to TR:G1184699 G1184699 TYROSYL-TRNA SYNTHETASE
;
ACCESSION AA086129
NID G1629697
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Carnivora; Homnidae; Homo.
REFERENCE 1 (bases 1 to 525)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E.,
Waterson,R., Williamson,A., Woldmann,P. and Wilson,R.
Washington, R., Williamson, A., Woldmann, P. and Wilson, R.
WashU-Merck Est Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT
Contact: Wilson RK
WashU-Merck Est Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 330.
Location/Qualifiers
1. 525
/organism="Homo sapiens"
/notes="Organ: colon; Vector: pBluescript SK-; Site:1:
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:
Oligo dT: T-84 colonic epithelial cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTT 3'"
/db_xref="taxon:9606"
/clone="511328"
/clone_lib="Strataene colon (#937204)"
/lab_host="SOLR cells (kanamycin resistant)"
<1. >525
BASE COUNT 144 a 125 c 149 g 105 t 2 others
ORIGIN

Query Match 54.6%; Score 275; DB 15; Length 525;
Best Local Similarity 96.4%; Pred. No. 0.00e+00;
Matches 322; Conservative 0; Mismatches 5; Indels 7; Gaps 7;

Db 155 GAGGAGGTATCCATCCCGCGGTGAGATATCCGTGTGGGGAAAAATCATCAGCTGTGAGAG 214
|||||
QY 94 GAGGAGGTATCCATCCCGCGGTGAGATATCCGTGTGGGGAAAAATCATCAGCTGTGAGAG 153
|||||
Db 215 CACCCAGATCACACAGACCGCTGATGTAGAGAGATTACGTTGGGGGGAAGCGAACCACGG 274
|||||
QY 154 CACCCAGATCACACAGACCGCTGATGTAGAGAGATTACGTTGGGGGGAAGCGAACCACGG 213
|||||
Db 275 ACTGTGTGAAGCGCGCTGTACAGTTCGTGTCGCCAAGAGAGAAATTGCAGACAGCTGT 334
|||||
QY 214 ACTGTGTGA- GCGGCGCTGTACAGTTCGTGTCGCCAAGAGAGAAATTGCAGACAGCTGT 272
|||||
Db 335 AATGTCCTGTGTGCAACTGAACCCAGAAAGATGAGAGGAGTGAAGTCCCAAGCATGC 394
|||||

```

Query Match	53.0%	Score 267	DB 17	Length 337
Best Local Similarity	89.6%	Fred. No. 0.00e+00		
Matches 302	Conservative	0	Mismatches 35	Indels 0
			Gaps 0	

Db 1 GCACCCAGATGCAGATACCTTTATGTGAGAGAGATTGATGTGGGGAGACTGAACCCAG 60
 QY 153 GCACCCAGATGCAGACACCTGTATGTAGAGAGATTGACGTGGGGAGACTGAACCCAG 212
 Db 61 GACGTGTGTAGCGGGCCCTGTTACATTTGTGTGCCCAAGAGAACCTGCAGACAGCTGT 120
 QY 213 GACTGTGTGTAGCGGGCCCTGTTACATTTGTGTGCCCAAGAGAACCTGCAGACAGCTGT 272
 Db 121 GGTGTGTGTGTGCAATCTGAACCCAGAGATGAGAGCGGTGAGCTGCAGAGCGATGT 180
 QY 273 AGTGTGTGTGTGCAACCTGTAACCCAGAGATGAGAGAGTGTGAGTCCAGAGCGATGT 332
 Db 181 ACTGTGTCTTCTGTAGAGGGGTGAGCGCCAGTTGAACCTCTGAGACCTCTGTCTGT 240
 QY 333 TCTGTGTCTTCTATATAGAGAGGATTAACCGCAGGTGTAACCTCTGAGACCTCTGCGGAG 392
 Db 241 CTCGTCTCTGTGTGACAGAGTGTGTGTACAGGCGTATGAGAGGCGCCAGCATGAGGA 300
 QY 393 CTCGTCTCTGTGTGAGCAGGTGTGTGTGAGAGGCGTATGAGAGGCGCCAGCATGAGGA 452
 Db 301 GCTCAAGCCCAAGAGAGAGTCTTGTGAGAGGCTGCAG 337
 QY 453 GCTCAAGCCCAAGAGAGAGTCTTGTGAGAGGCTGCAG 489

RESULT 7
 LOCUS N87033 243 bp mRNA EST 01-APR-1996
 DEFINITION L1914F fetal heart, lambda ZAP Express Homo sapiens cDNA clone
 L1914.5' similar to EST(Y625D10.R1 P23395 METHIONYL-TRNA
 SYNTHETASE).

ACCESSION N87033
 NID g1440235
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Euthera; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 243)
 Llew.C.C.
 CDNAS from fetal heart (1996)
 JOURNAL Unpublished (1996)
 COMMENT

CONTACT: Llew CC
 Molecular Cardiology
 University of Toronto
 Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
 Tel: 4169788758
 Fax: 4169785650
 Email: llewcc@utcc.utoronto.ca
 Seq primer: GAATTTAACTCTCACTAAAGCG.
 Location/Qualifiers

FEATURES
 source
 1.243
 /organism="Homo sapiens"
 /note="Vector: Lambda ZAP Express; site_1: EcoRI; site_2:
 /xhoI; mRNA was purified from human fetal hearts (8-10
 weeks). cDNA was synthesized using a XhoI-oligo dT
 adaptor-primer. EcoRI adaptors were ligated, followed by
 digestion with XhoI for directional cloning into
 preligated lambda ZAP Express."
 /db_xref="taxon:9606"
 /clone="L1914"
 /clone_1db="Fetal heart, Lambda ZAP Express"
 /lab_host="E. coli XL1-Blue"
 <1..>243

BASE COUNT 57 a 58 c 77 g 51 t
 ORIGIN
 mRNA
 Query Match 43.5%; Score 219; DB 14; Length 243;
 Best Local Similarity 96.7%; Pred. No. 0.00e+00;
 Matches 237; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

Db 1 CCCAGATGCAGACGCTGTATGTAGAGAGATTGACGTGGGGAGACTGAACCCAGCGAG 60

QY 156
 Db 61
 QY 216
 Db 121
 QY 276
 Db 180
 QY 336
 Db 239
 QY 396

|||||
 CCAGATGCAGACACCTGTATGTAGAGAGATTGACGTGGGGAGACTGAACCCAGCGAG 215
 TGTGTGTAGCGGCGCTGTGTACAGTGTGTGTGCCCAAGAGAGAACTGCAGACAGCTGTAGT 120
 TGTGTGTAGCGGCGCTGTGTACAGTGTGTGTGCCCAAGAGAGAACTGCAGACAGCTGTAGT 275
 TGTGTGTGTGTGCAATCTGAACCCAGAGATGAGAGAGTGTGAGTCCAGAGCGATGTCT 335
 GGTGTGTGTGTGCAACCTGTAACCCAGAGATGAGAGAGTGTGAGTCCAGAGCGATGTCT 395
 GTGTCTTCTATATAGAGAGGATTA-CCGCGAGGTGTAACCTCTGTGCGCTCGGAGGTTTC 238
 GTGTCTTCTATATAGAGAGGATTAACCGCAGGTGTAACCTCTGTGAGACCTCTGCGGAGGCTC 395
 TGCTC 243
 TGTCTC 400

RESULT 8
 LOCUS AA003398 423 bp mRNA EST 19-JUL-1996
 DEFINITION mg57b09.t1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
 clone 437081 5' similar to PIR:A55053 A55053
 endothelial monocyte-activating protein II precursor - mouse ; .

ACCESSION AA003398
 NID g1446863
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Euthera; Rodentia; Sciurognathi; Myomorpha; Muridae;
 Murinae; Mus.
 1 (bases 1 to 423)
 Marra,M., Hallier,U., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Stepcio,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The Washu-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT

CONTACT: Marra M/Mouse EST Project
 Washu-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LML; contact the
 IMAGE Consortium (info@image.lml.gov) for further information.
 MGI:262417
 Seq primer: ETPprimer
 High quality sequence stop: 343.
 Location/Qualifiers

FEATURES
 source
 1.423
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; site_1: Not I; site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGACAGGAGGCGGCGGAGAAATTTTGTGTGTGTGTGTGTGTGT
 T 3']; on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
 State Univ., from 2 j]; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M.Fatima Bonaldo."
 /db_xref="taxon:10090"
 /clone="437081"
 /clone_1db="Soares mouse embryo NbME13.5 14.5"

			/sex="unknown"	
			/tissue_type="embryo"	
			/dev_stage="13.5-14.5dpc total fetus"	
			/lab_host="DH10B"	
			<1..->423	
MRNA				
BASE COUNT	98 a	115 c	120 g	90 t
ORIGIN				
Query Match	42.7%	Score 215;	DB 14;	Length 423;
Best Local Similarity 90.3%;	Pred. No. 0.00e+00;			
Matches 241;	Conservative 0;	Mismatches 26;	Indels 0;	Gaps 0;
Db	157	GAACTGCAGGACAGCGCTGGTGGTGCTGTGTCATTCGAATCCCAAGAATGAGAGGC	216	
QY	253	GAACTGCAGGACAGCGCTGGTGGTGCTGTGTCATTCGAATCCCAAGAATGAGAGGA	312	
Db	217	GTGAGCTGGCAGGGGATGCTACTGTGTGTGTTGTGTGTAAGAGGGGTAGCCGCCAGGTGAA	276	
QY	313	GTCCAGTCCCAAGGAGCATGCTGTGTGTGTTGTATTAAGAGGATATAAACCCGCAAGGTGAA	372	
Db	277	CCTCTGAGACCCTCCTGCTGCTCTGCTCTCCTCGGTGGAACGATGTTGACAGGCTATGAG	336	
QY	373	CCTCTGAGACCCTCCTGCTGCTCTGCTCTCCTCGGTGGAACGATGTTGATGAAGGCTATGAA	432	
Db	337	AAGGGCCAGCCAGATGAGGAGCTCAAGCCCAAGAAAGATTGTTGAGAAGCTGCAGGCC	396	
QY	433	AAGGGCCAGCCAGATGAGGAGAGCTCAAGCCCAAGAAAGATTGTTGAGAAGTTGCAGGCT	492	
Db	397	GACTTTAAATTTCTGAGAGATGCATC	423	
QY	493	GACTTCAAATTTCTGAGAGATGCATC	519	
RESULT LOCUS	9	AA221081	517 bp	EST
DEFINITION		aw65f03.r1 Soares mouse JMWEL2	5 Mus musculus	CDNA clone 659933 5'
ACCESSION		AA221081		
NID		G1840276		
KEYWORDS		EST.		
SOURCE		house mouse.		
ORGANISM		Mus musculus		
REFERENCE		Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;		
AUTHORS		Mus.		
TITLE		1 (bases 1 to 517)		
JOURNAL		Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,		
COMMENT		Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,		
		Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,		
		Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and		
		Waterston,R.		
		The Mashu-HMI Mouse EST Project		
		Unpublished (1996)		
		Contact: Marra M/Mouse EST Project		
		Mashu-HMI Mouse EST Project		
		Washington University School of Medicine		
		444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
		Tel: 314 286 1800		
		Fax: 314 286 1810		
		Email: mousese@watson.wustl.edu		
		This clone is available royalty-free through LML ; contact the		
		IMAGE Consortium (infoimage.llnl.gov) for further information.		
		MG:405781		
		Seq primer: -28ml3 rev2 ET from Amersham		
		High quality sequence stop: 493.		
FEATURES		Location/Qualifiers		
SOURCE		1..517		

REFERENCE	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	BASE COUNT	ORIGIN
ADAMS, M.D., KERLAVAGE, A.R., FLEISCHMANN, R.D., FULNER, R.A., BULL, C.J., DEE, N.H., KIRKNESS, E.F., WEINSTOCK, K.G., GOCAYNE, J.D., WHITE, O., SUTTON, G., BLAKE, J.A., BRANDON, R.C., MAN-WEI, C., CLAYTON, R.A., CLINE, T.R., COTTON, M.D., EARLE-HUGHES, J., FINE, L.D., FITZGERALD, L.M., FITZHUGH, W.M., FRITCHMAN, J.L., GEOGRAPHOS, N.S., GLODER, A., GNEHM, C.L., HANNA, M.C., HEDBLOM, E., HINKLE, P.S., JR., KELLEY, J.W., KELLEY, J.C., LIU, L.-I., MAMMATROS, S.M., MERRICK, J.M., MORENO-PALANQUES, R.F., McDONALD, L.A., NGUYEN, D.T., PELLIGRINO, S.M., PHILLIPS, C.A., RYDER, S.E., SCOTT, J.L., SAUDEK, D.M., SHIRLEY, R., SMALL, K.V., SPRIGGS, T.A., UTERBACK, T.R., WEIDMAN, J.F., LI, Y., BEHNATIK, D.P., CAO, L., CEPEDA, M.A., COLEMAN, T.A., COLLINS, E.J., DINKEL, D., FENG, D.-F., FERTLE, A., FISCHER, C., HASTINGS, G.A., HE, W.W., HU, J.S., GREENE, J.M., GRUBER, J., HUDSON, P., KIM, A.K., KOZAK, D.L., KUNSCHE, C., HUNGJUN, J., LI, H., WEISSNER, P.S., OLSEN, H., RAYMOND, L., WEI, Y.F., WING, J., XU, C., YU, G.L., RUBEN, S.M., DILLON, P.J., FANNON, M.R., ROSEN, C.A., HASSETLINE, W.A., FIELDS, C., FRASER, C.M., and VENTER, J.C.	AA327316	EST306528	AA327316	g1979561	EST.	human.	Homo sapiens	138 a	123 c
ADAMS, M.D., KERLAVAGE, A.R., FLEISCHMANN, R.D., FULNER, R.A., BULL, C.J., DEE, N.H., KIRKNESS, E.F., WEINSTOCK, K.G., GOCAYNE, J.D., WHITE, O., SUTTON, G., BLAKE, J.A., BRANDON, R.C., MAN-WEI, C., CLAYTON, R.A., CLINE, T.R., COTTON, M.D., EARLE-HUGHES, J., FINE, L.D., FITZGERALD, L.M., FITZHUGH, W.M., FRITCHMAN, J.L., GEOGRAPHOS, N.S., GLODER, A., GNEHM, C.L., HANNA, M.C., HEDBLOM, E., HINKLE, P.S., JR., KELLEY, J.W., KELLEY, J.C., LIU, L.-I., MAMMATROS, S.M., MERRICK, J.M., MORENO-PALANQUES, R.F., McDONALD, L.A., NGUYEN, D.T., PELLIGRINO, S.M., PHILLIPS, C.A., RYDER, S.E., SCOTT, J.L., SAUDEK, D.M., SHIRLEY, R., SMALL, K.V., SPRIGGS, T.A., UTERBACK, T.R., WEIDMAN, J.F., LI, Y., BEHNATIK, D.P., CAO, L., CEPEDA, M.A., COLEMAN, T.A., COLLINS, E.J., DINKEL, D., FENG, D.-F., FERTLE, A., FISCHER, C., HASTINGS, G.A., HE, W.W., HU, J.S., GREENE, J.M., GRUBER, J., HUDSON, P., KIM, A.K., KOZAK, D.L., KUNSCHE, C., HUNGJUN, J., LI, H., WEISSNER, P.S., OLSEN, H., RAYMOND, L., WEI, Y.F., WING, J., XU, C., YU, G.L., RUBEN, S.M., DILLON, P.J., FANNON, M.R., ROSEN, C.A., HASSETLINE, W.A., FIELDS, C., FRASER, C.M., and VENTER, J.C.	AA327316	EST306528	AA327316	g1979561	EST.	human.	Homo sapiens	138 a	123 c
ADAMS, M.D., KERLAVAGE, A.R., FLEISCHMANN, R.D., FULNER, R.A., BULL, C.J., DEE, N.H., KIRKNESS, E.F., WEINSTOCK, K.G., GOCAYNE, J.D., WHITE, O., SUTTON, G., BLAKE, J.A., BRANDON, R.C., MAN-WEI, C., CLAYTON, R.A., CLINE, T.R., COTTON, M.D., EARLE-HUGHES, J., FINE, L.D., FITZGERALD, L.M., FITZHUGH, W.M., FRITCHMAN, J.L., GEOGRAPHOS, N.S., GLODER, A., GNEHM, C.L., HANNA, M.C., HEDBLOM, E., HINKLE, P.S., JR., KELLEY, J.W., KELLEY, J.C., LIU, L.-I., MAMMATROS, S.M., MERRICK, J.M., MORENO-PALANQUES, R.F., McDONALD, L.A., NGUYEN, D.T., PELLIGRINO, S.M., PHILLIPS, C.A., RYDER, S.E., SCOTT, J.L., SAUDEK, D.M., SHIRLEY, R., SMALL, K.V., SPRIGGS, T.A., UTERBACK, T.R., WEIDMAN, J.F., LI, Y., BEHNATIK, D.P., CAO, L., CEPEDA, M.A., COLEMAN, T.A., COLLINS, E.J., DINKEL, D., FENG, D.-F., FERTLE, A., FISCHER, C., HASTINGS, G.A., HE, W.W., HU, J.S., GREENE, J.M., GRUBER, J., HUDSON, P., KIM, A.K., KOZAK, D.L., KUNSCHE, C., HUNGJUN, J., LI, H., WEISSNER, P.S., OLSEN, H., RAYMOND, L., WEI, Y.F., WING, J., XU, C., YU, G.L., RUBEN, S.M., DILLON, P.J., FANNON, M.R., ROSEN, C.A., HASSETLINE, W.A., FIELDS, C., FRASER, C.M., and VENTER, J.C.	AA327316	EST306528	AA327316	g1979561	EST.	human.	Homo sapiens	138 a	123 c
ADAMS, M.D., KERLAVAGE, A.R., FLEISCHMANN, R.D., FULNER, R.A., BULL, C.J., DEE, N.H., KIRKNESS, E.F., WEINSTOCK, K.G., GOCAYNE, J.D., WHITE, O., SUTTON, G., BLAKE, J.A., BRANDON, R.C., MAN-WEI, C., CLAYTON, R.A., CLINE, T.R., COTTON, M.D., EARLE-HUGHES, J., FINE, L.D., FITZGERALD, L.M., FITZHUGH, W.M., FRITCHMAN, J.L., GEOGRAPHOS, N.S., GLODER, A., GNEHM, C.L., HANNA, M.C., HEDBLOM, E., HINKLE, P.S., JR., KELLEY, J.W., KELLEY, J.C., LIU, L.-I., MAMMATROS, S.M., MERRICK, J.M., MORENO-PALANQUES, R.F., McDONALD, L.A., NGUYEN, D.T., PELLIGRINO, S.M., PHILLIPS, C.A., RYDER, S.E., SCOTT, J.L., SAUDEK, D.M., SHIRLEY, R., SMALL, K.V., SPRIGGS, T.A., UTERBACK, T.R., WEIDMAN, J.F., LI, Y., BEHNATIK, D.P., CAO, L., CEPEDA, M.A., COLEMAN, T.A., COLLINS, E.J., DINKEL, D., FENG, D.-F., FERTLE, A., FISCHER, C., HASTINGS, G.A., HE, W.W., HU, J.S., GREENE, J.M., GRUBER, J., HUDSON, P., KIM, A.K., KOZAK, D.L., KUNSCHE, C., HUNGJUN, J., LI, H., WEISSNER, P.S., OLSEN, H., RAYMOND, L., WEI, Y.F., WING, J., XU, C., YU, G.L., RUBEN, S.M., DILLON, P.J., FANNON, M.R., ROSEN, C.A., HASSETLINE, W.A., FIELDS, C., FRASER, C.M., and VENTER, J.C.	AA327316	EST306528	AA327316	g1979561	EST.	human.	Homo sapiens	138 a	123 c
ADAMS, M.D., KERLAVAGE, A.R., FLEISCHMANN, R.D., FULNER, R.A., BULL, C.J., DEE, N.H., KIRKNESS, E.F., WEINSTOCK, K.G., GOCAYNE, J.D., WHITE, O., SUTTON, G., BLAKE, J.A., BRANDON, R.C., MAN-WEI, C., CLAYTON, R.A., CLINE, T.R., COTTON, M.D., EARLE-HUGHES, J., FINE, L.D., FITZGERALD, L.M., FITZHUGH, W.M., FRITCHMAN, J.L., GEOGRAPHOS, N.S., GLODER, A., GNEHM, C.L., HANNA, M.C., HEDBLOM, E., HINKLE, P.S., JR., KELLEY, J.W., KELLEY, J.C., LIU, L.-I., MAMMATROS, S.M., MERRICK, J.M., MORENO-PALANQUES, R.F., McDONALD, L.A., NGUYEN, D.T., PELLIGRINO, S.M., PHILLIPS, C.A., RYDER, S.E., SCOTT, J.L., SAUDEK, D.M., SHIRLEY, R., SMALL, K.V., SPRIGGS, T.A., UTERBACK, T.R., WEIDMAN, J.F., LI, Y., BEHNATIK, D.P., CAO, L., CEPEDA, M.A., COLEMAN, T.A., COLLINS, E.J., DINKEL, D., FENG, D.-F., FERTLE, A., FISCHER, C., HASTINGS, G.A., HE, W.W., HU, J.S., GREENE, J.M., GRUBER, J., HUDSON, P., KIM, A.K., KOZAK, D.L., KUNSCHE, C., HUNGJUN, J., LI, H., WEISSNER, P.S., OLSEN, H., RAYMOND, L., WEI, Y.F., WING, J., XU, C., YU, G.L., RUBEN, S.M., DILLON, P.J., FANNON, M.R., ROSEN, C.A., HASSETLINE, W.A., FIELDS, C., FRASER, C.M., and VENTER, J.C.	AA327316	EST306528	AA327316	g1979561	EST.	human.	Homo sapiens	138 a	123 c
ADAMS, M.D., KERLAVAGE, A.R., FLEISCHMANN, R.D., FULNER, R.A., BULL, C.J., DEE, N.H., KIRKNESS, E.F., WEINSTOCK, K.G., GOCAYNE, J.D., WHITE, O., SUTTON, G., BLAKE, J.A., BRANDON, R.C., MAN-WEI, C., CLAYTON, R.A., CLINE, T.R., COTTON, M.D., EARLE-HUGHES, J., FINE, L.D., FITZGERALD, L.M., FITZHUGH, W.M., FRITCHMAN, J.L., GEOGRAPHOS, N.S., GLODER, A., GNEHM, C.L., HANNA, M.C., HEDBLOM, E., HINKLE, P.S., JR., KELLEY, J.W., KELLEY, J.C., LIU, L.-I., MAMMATROS, S.M., MERRICK, J.M., MORENO-PALANQUES, R.F., McDONALD, L.A., NGUYEN, D.T., PELLIGRINO, S.M., PHILLIPS, C.A., RYDER, S.E., SCOTT, J.L., SAUDEK, D.M., SHIRLEY, R., SMALL, K.V., SPRIGGS, T.A., UTERBACK, T.R., WEIDMAN, J.F., LI, Y., BEHNATIK, D.P., CAO, L., CEPEDA, M.A., COLEMAN, T.A., COLLINS, E.J., DINKEL, D., FENG, D.-F., FERTLE, A., FISCHER, C., HASTINGS, G.A., HE, W.W., HU, J.S., GREENE, J.M., GRUBER, J., HUDSON, P., KIM, A.K., KOZAK, D.L., KUNSCHE, C., HUNGJUN, J., LI, H., WEISSNER, P.S., OLSEN, H., RAYMOND, L., WEI, Y.F., WING, J., XU, C., YU, G.L., RUBEN, S.M., DILLON, P.J., FANNON, M.R., ROSEN, C.A., HASSETLINE, W.A., FIELDS, C., FRASER, C.M., and VENTER, J.C.	AA327316	EST306528	AA327316	g1979561	EST.	human.	Homo sapiens	138 a	123 c
ADAMS, M.D., KERLAVAGE, A.R., FLEISCHMANN, R.D., FULNER, R.A., BULL, C.J., DEE, N.H., KIRKNESS, E.F., WEINSTOCK, K.G., GOCAYNE, J.D., WHITE, O., SUTTON, G., BLAKE, J.A., BRANDON, R.C., MAN-WEI, C., CLAYTON, R.A., CLINE, T.R									

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

High quality sequence stop: 468.

FEATURES

Location/Qualifiers
1. 500
/organism="Mus musculus"

/strain="C57BL/6J"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAGAGCGGCGCGGAGAAATTTTCTTTTCTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos (total RNA provided by Minoru Ko, Wayne State Univ., from 2 1; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT733 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

mRNA 130 a 133 c 130 g 107 t

ORIGIN

Query Match 31.2%; Score 157; DB 14; Length 500;
Best Local Similarity 89.8%; Pred. No. 3.33e-263;

Matches 177; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 1 CTGCTGAACGAGTCTTGTACAGGCGCTATGAGAGGCGCCAGCAGATGAGAGCTCAACG 60

QY 401 CTGCTGAACGAGTCTTGTACAGGCGCTATGAGAGGCGCCAGCAGATGAGAGCTCAACG 460

Db 61 CCAAGAGAAAGTCTTGTAGAGAGCTGCGAGCGGCGACTTTAAATTTCTGAGAGTGCATCG 120

QY 461 CCAAGAGAAAGTCTTGTAGAGAGCTGCGAGCGGCGACTTTAAATTTCTGAGAGTGCATCG 520

Db 121 CACAGTGAACCAACCAACTTCATGACCAAGCTGGGATTCCTCCCTGTAATCACTAA 180

QY 521 CACAGTGAACCAACCAACTTCATGACCAAGCTGGGATTCCTCCCTGTAATCACTAA 580

Db 181 AAGGGGGTAACATCAGC 197

QY 581 AAGGGGGTAACATCAGC 597

RESULT 13

LOCUS AA051121 510 bp mRNA EST 09-SEP-1996

DEFINITION m74c08.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA

ACCESSION AA051121

NID 61530793

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 510)

AUTHORS Marra,M., Hallier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

TITLE

JOURNAL The WashU-HMI Mouse EST Project

COMMENT Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

FEATURES

Location/Qualifiers
1. 510
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAGAGCGGCGCGGAGAAATTTTCTTTTCTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos (total RNA provided by Minoru Ko, Wayne State Univ., from 2 1; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT733 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 495.

Location/Qualifiers

1. 510

/organism="Mus musculus"

/strain="C57BL/6J"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAGAGCGGCGCGGAGAAATTTTCTTTTCTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos (total RNA provided by Minoru Ko, Wayne State Univ., from 2 1; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT733 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

/db_xref="taxon:10090"

/clone.lib="Soares mouse embryo NBME13.5 14.5"

/sex="unknown"

/tissue="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

/lab_host="DH10B"

BASE COUNT 134 a 132 c 130 g 114 t

ORIGIN

Query Match 31.2%; Score 157; DB 14; Length 510;
Best Local Similarity 89.8%; Pred. No. 3.33e-263;

Matches 177; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 1 CTGCTGAACGAGTCTTGTACAGGCGCTATGAGAGGCGCCAGCAGATGAGAGCTCAACG 60

QY 401 CTGCTGAACGAGTCTTGTACAGGCGCTATGAGAGGCGCCAGCAGATGAGAGCTCAACG 460

Db 61 CCAAGAGAAAGTCTTGTAGAGAGCTGCGAGCGGCGACTTTAAATTTCTGAGAGTGCATCG 120

QY 461 CCAAGAGAAAGTCTTGTAGAGAGCTGCGAGCGGCGACTTTAAATTTCTGAGAGTGCATCG 520

Db 121 CACAGTGAACCAACCAACTTCATGACCAAGCTGGGATTCCTCCCTGTAATCACTAA 180

QY 521 CACAGTGAACCAACCAACTTCATGACCAAGCTGGGATTCCTCCCTGTAATCACTAA 580

Db 181 AAGGGGGTAACATCAGC 197

QY 581 AAGGGGGTAACATCAGC 597

RESULT 14

LOCUS H13315 437 bp mRNA EST 27-JUN-1995

DEFINITION y172007.r1 Homo sapiens cDNA clone 43366 5' similar to

SP:SYC_YEAST P36421 TYROSYL-TRNA SYNTHETASE, CYTOPLASMIC ;

ACCESSION H13315

NID 9878135

KEYWORDS EST.

SOURCE human clone=43366 library=Soares infant brain INB vector=Lafind BA

host=DH10B (ampicillin resistant) primer=H13P1 RstIel-Not I

RstIel2-Hind III whole brain from a 73 days post natal female. 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5',

AAGTGAAGAAATTCGCGCGCGGAGAAATTTTCTTTTCTTTTCTTTT

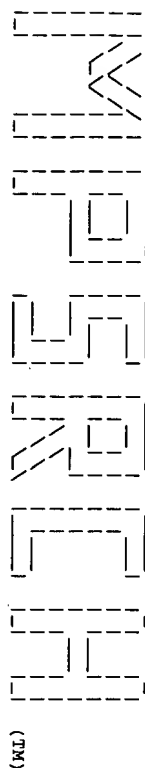
T 3'], double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with

Not I and directionally cloned into the Not I and Hind III sites of

the Lafind BA vector. Library went through one round of

normalization. Library constructed by Bento Soares and M. Fatima

Bonaldo.



(7M)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msearch_nu n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Apr 24 08:13:01 1998; Maspar time 430.29 Seconds

Tabular output not generated. 1384.997 Million cell updates/sec

Title: >US-08-483-534A-1
Description: (94-597) from US08483534A.seq (1 of 2)
Perfect Score: 504
N.A. Sequence: 94 GAGGAGGTGATCCCATCCG.....TGAAGGGGGAGACATTAGC 597
Comp: CTCCTCCAGTAGGATAGGCG.....ACTTCCCTTCCTGTAATCG

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 354530 segs, 591221332 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb153

1:em_in 2:em_or 3:em_com 4:em_ov 5:em_pl 6:em_htg
7:em_hum1 8:em_hum2 9:em_da 10:em_ro 11:em_un 12:em_vi
13:em_pat
genbank105

Database: 14:gb_ro 15:gb_com 16:gb_ov 17:gb_in 18:gb_pl 19:gb_da
20:gb_st 21:gb_vi 22:gb_ph 23:gb_sy 24:gb_un 25:gb_pat
26:gb_htg

Statistics: Mean 10.143; Variance 5.076; scale 1.998

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
C 1	82	16.3	21948	17	CEFS8B3	9.54e-41
C 2	82	16.3	85953	26	CEH04L24	9.54e-41
C 3	82	16.3	131318	26	CEY42E10	9.54e-41
C 4	81	16.1	498	23	ASEMAPIT	9.54e-41
C 5	73	14.5	1068	14	NMDU10118	6.67e-34
C 6	73	14.5	1068	14	Mus musculus endocheli	6.67e-34
C 7	72	14.3	7218	25	150092	6.67e-34
C 8	66	13.1	1254	14	AF021800	1.15e-28
C 9	33	6.5	1703	18	SCYGL105W	9.53e-06
C 10	33	6.5	2473	24	SCARCI	9.53e-06
C 11	33	6.5	12325	18	SCVILIEFT	9.53e-06
C 12	32	6.3	1745	18	SCU31348	3.93e-05
C 13	31	6.2	215	25	128278	1.59e-04
C 14	30	6.0	11157	19	ECAE000389	6.34e-04

Result	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
1	CEFS8B3	21948 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
2	CEH04L24	85953 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
3	CEY42E10	498 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
4	ASEMAPIT	498 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
5	NMDU10118	1068 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
6	Mus musculus endocheli	1068 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
7	150092	7218 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
8	AF021800	1254 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
9	SCYGL105W	1703 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
10	SCARCI	2473 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
11	SCVILIEFT	12325 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
12	SCU31348	1745 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
13	128278	215 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		
14	ECAE000389	11157 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.		

ALIGNMENTS

1
CEFS8B3 21948 bp DNA INV 23-JAN-1998
Caenorhabditis elegans cosmid F58B3, complete sequence.
g1370027
HTG: Arabidopsis DNA-damage-repair D11 protein like;
interferon-related PC4 protein like; Methionyl-tRNA synthetase.
Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryotae; Metazoa; Nematoda; Secernentea; Rhabditiida; Rhabditiida;
Rhabditiida; Rhabditiida; Rhabditiida; Peloderiinae; Caenorhabditis.
1 (bases 1 to 21948)
Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Garner, A., Green, P., Hawkins, J., Hillier, L., Jier, M.,
Johnson, L., Jones, M., Kersey, J., Kistner, J., Laister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurtry, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rifen, L., Roop, A.,
Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonnenhammer, E.,
Straden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaubin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, V., and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
Nature in press
2 (bases 1 to 21948)
Harris, B.
Direct Submission
Submitted (13-MAY-1996) Louis, MO 63110, USA. E-mail:
jesseanger.ac.uk or twenematecode.wustl.edu
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note. IMPORTANT: This
sequence is NOT necessarily the entire insert of clone F58B3. It
may be shorter because we only sequence overlapping sections once.

LOCUS	4	ASEMAP11	498 bp	DNA	SYN	10-OCT-1997
DEFINITION	Artificial sequence DNA for cytokine EMAP11.					
ACCESSION	Y08737					
NID	91619591					
KEYWORDS	EMAP11 gene.					
SOURCE	synthetic construct.					
ORGANISM	artificial sequence.					
REFERENCE	1 (bases 1 to 498)					
AUTHORS	Schlusener,H.J., Seid,K., Zhao,Y. and Meyermann,R.					
TITLE	Localization of endothelial-monocyte-activating polypeptide II (EMAP II), a novel proinflammatory cytokine, to lesions of experimental autoimmune encephalomyelitis, neuritis and uveitis: expression by monocytes and activated microglial cells					
JOURNAL	Glia 20 (4), 365-372 (1997)					
MEDLINE	97404153					
REFERENCE	2 (bases 1 to 498)					
AUTHORS	Schlusener,H.J.					
TITLE	Direct Submission					
JOURNAL	Submitted (01-OCT-1996) H.J. Schlusener, Institute of Brain Research, Calver Str. 3, D-72076 Tuebingen, FRG					
FEATURES	Location/Qualifiers					
source	1..498					
gene	/organism="synthetic construct"					
	/db_xref="taxon:32630"					
	1..498					
CDS	/gene="human EMAP11"					
	<1..>498					
	/gene="human EMAP11"					
	/codon_start=1					
	/db_xref="PID:e274145"					
	/db_xref="PID:g1619592"					
	/transl_table=11					
	/translation="SKPIDVSRDLRIGCIITANKHPDADSLYEEVDVEIARIVV SGLVNHYPLEQMQRNRYILLCNLKPARMRGVLQAMVWCASSPERIKIILAPNPSVPG DRIIFDPAPEPDKENLPMKRKIMEOIOPDLHTNDECAVATKYGFVEVKGQVCHACTM SNSGIK"					
BASE COUNT	130 a	111 c	147 g	109 t	1 others	
ORIGIN						
Query Match	16.1%; Score 81; DB 23; Length 498;					
Best Local Similarity	67.5%; Pred. No. 5.56e-40;					
Matches 156; Conservative	0;	Mismatches 75;	Indels 0;	Gaps 0;		
Db	21	CCGCTGGATCGCTATTGGCTGCATTATTAACGCGGGTAACACCGCGGATGCAGATAG	80			
QY	111	CCGGCTGGATTCCTGTGGGGAAATATATCTCTGTGGAGACACCCAGATGCAGACAG	170			
Db	81	CCTGTATGTGAAGAAGTGAATGTGGTGAATTGCGCGCTACCGTGTCTGTGGCCT	140			
QY	171	CCTGTATGTGAAGAAGTGAATGTGGTGAATTGCGCGCTACCGTGTCTGTGGCCT	230			
Db	141	GCTGACCATGTGCGCGCTGGAACACATGCAGAACCGTATGGTGTCTGTGTGCACCT	200			
QY	231	GCTGACCATGTGCGCGCGCTGGAACACATGCAGAACCGTATGGTGTCTGTGTGCACCT	290			

DB	201	GAACACGGCGAAATGCGTGGCGGCTCTCAGGCAATGGTATGTCGCG	251
QY	291	GAACACCGAAGATGAGAGAGTCGACTGCCAAGCGATGCTCTGTCG	341
RESULT	5		
LOCUS	MMU10118	1068 bp	mRNA
DEFINITION	Mus musculus endothelial-monocyte activating polypeptide II mRNA,		
ACCESSION	U10118		
NID	9498911		
KEYWORDS			
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryotes: mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sclurognathi; Myomorphia; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1068) Kao,J., Houck,K., Fan,Y., Haehnel,I., Libutti,S.K., Kayton,M.L., Gritschelt,T., Chabot,J., Noygrod,R., Greenberg,S., Kuang,W.-J., Leung,D.W., Hayward,J.R., Kistel,W., Heath,M., Brett,J. and Stern,D.M.		
TITLE	Characterization of a novel tumor-derived cytokine.		
JOURNAL	Endothelial-monocyte activating polypeptide II		
MEDLINE	J Biol. Chem. 269 (40), 25106-25119 (1994)		
REFERENCE	95014290		
AUTHORS	2 (bases 1 to 1068) Kao,J., Fan,Y., Haehnel,I., Brett,J., Greenberg,S., Claus,M., Kayton,M., Houck,K., Kistel,W., Seljelid,R., Burnier,J. and Stern,D.		
TITLE	A peptide derived from the amino terminus of endothelial-monocyte-activating polypeptide II modulates mononuclear and polymorphonuclear leukocyte functions, defines an apparently novel cellular interaction site, and induces an acute inflammatory response		
JOURNAL	J. Biol. Chem. 269 (13), 9774-9782 (1994)		
MEDLINE	94193655		
REFERENCE	3 (bases 1 to 1068)		
AUTHORS	Kao,J., Ryan,J., Brett,G., Chen,J., Shen,H., Fan,Y., Godman,G., Fanilletti,P.C., Wang,F., Pan,Y.E., Stern,D. and Claus,M.		
TITLE	Endothelial monocyte-activating polypeptide II. A novel tumor-derived polypeptide that activates host-response mechanisms		
JOURNAL	J. Biol. Chem. 267 (28), 20239-20247 (1992)		
MEDLINE	93013897		
REFERENCE	4 (bases 1 to 1068)		
AUTHORS	Houck,K.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (27-MAY-1994)		
MEDLINE	Pharmaceuticals Corp., P.O. Box 52350, Durham, NC 27711, USA		
REFERENCE	Location/Qualifiers		
AUTHORS	1. 1068		
TITLE	/organism="Mus musculus"		
JOURNAL	/db.xref="taxon:10090"		
MEDLINE	/cell_line="Meth A"		
REFERENCE	/cell_type="fibrosarcoma"		
AUTHORS	1. 1068		
TITLE	64. 996		
JOURNAL	/standard_name="EMAP II"		
MEDLINE	/codon_start=1		
REFERENCE	/product="endothelial-monocyte activating polypeptide II"		
AUTHORS	/db.xref="PID:9498912"		
TITLE	/translation="MATNDVAVLRLEDKGAADQIIIEYLKQVALLKERAILOATNRE		
JOURNAL	EKRIKRVNALKKEIEIKELKQELILAEIHNVEQVRALSPLOTNCTASVSQSPSV		
MEDLINE	ATVAPSPKQIKAGEEKYKTEKTEKGEKKEKQSAASTDSKPIDASRLDIRICI		
REFERENCE	VTKKHPKPADSLYEEVDGCEAKPRVYSGLNVHVPQGNQNMVYLCTLRKARKG		
AUTHORS	VLSQAMVCASSPEKVEITLAPRPGSVPGDITITDAPGEDEKELNPKKIKWEDIQDL		
TITLE	HTNAECVATYKGAPEYKGGVGRQDTMANSGLK"		
JOURNAL	496. 993		
MEDLINE	/standard_name="EMAP II"		
REFERENCE	/product="endothelial-monocyte activating polypeptide II"		
AUTHORS	1037. 1042		
TITLE	mat_peptide		
JOURNAL	polya_signal		
MEDLINE	328 a		
REFERENCE	212 c		
AUTHORS	297 g		
TITLE	231 t		
JOURNAL	BASE COUNT		
MEDLINE	ORIGIN		


```

/obj.xref="PID:92754814"
/translation="MCEYFRRLPGTAPBSSPPAPATHRLRLGLGELRYRMRMTCRTF
AKATNDVNLKRLQKGEADQIIETELKQVALLKRAVLDAITLREKKLKEVENAKLKE
KEIHELQELQIAEIKONGVKQIPVYOSDIPADBSANVSISTSVSTISCSIKESB
SKGGEKKEVKEADIPDKKRGKKEKRLQSLAPADSKPVDYSIDRIGRIVYKKHPDA
DLVEEYEDVEAARPRVYISGLVNHVPLDQONMNVYLICLKRKKMGILSOAMVMO
ASSPEKEIILAPNGSVGDRIETFDAPFGEEDKELNPKKKIMEQIQDPLHNALCVAT
YKSPFEYKKGVCVRAQTMANSGIT"
BASE COUNT      400 a      251 c      343 g      260 t
ORIGIN
Query Match      13.1%; Score 66; DB 14; Length 1254;
Best Local Similarity 64.1%; Pred. No. 1.15e-28;
Matches 150; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
Db 681 TCTCGTCGTGATCTTCGATTTGGTGCATTTTCTACTGTCATAAAGCACCCGGATCGACAG 740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 109 TCCCGCGTGGANATCCGTGTGGGAAATCATCATCTGTGGAGAACACCCAGATGCACAGC 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 741 TCATCTGATAGTAGAAGAGTGGATGTGGGGGAGGACGCCCAAGAGCAGTCATCAGTGA 800
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 169 AGCGTGTATGATAGAGAAGTATGACGTGGGGGAGGTGAACCAAGCAGCTGTGTGAGCGGC 228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 801 CTGGTGAATCATGTTCCCTCTCTAGACCAAGATGCAGAAATCGAATGCTGATTTACTTGTAA 860
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 229 CTGTATCAGTTCGTGCTCCCAAGAGGAACTGCAGCAGAGGCTGTGATGGTGTCTGTGCAAC 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 861 CTGAAGCCCTGCAAAANATGGCGGATACTGTCTCAAGCAAGCTGATGCTGCT 914
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 289 CTGAATCCCAAGATGAGAGAGTGCAGTCCCAAGGAGCATCTTCTGTGCTCT 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

BASE COUNT	577 a	318 c	335 g	473 t	TINDGLEVIFKDEERDHPVRKLTNNKGSEFKVASIANNQVR*
ORIGIN					
Query Match			6.5%;	Score 33;	DB 18; Length 1703;
Best Local Similarity			59.8%;	Pred. No. 9.53e-06;	
Matches 146; Conservative			0;	Mismatches 95;	Indels 3; Gaps 1;
Db 1043	CCATCCGCAATAGATTTCGGTGTGTTTATTCAAAAAGCCATCAACACCCAGATGCC	1102			
Qy 106	CCATCCCGCGCGATATCCGTGTGGGAAAATCATCTGCTGGAGAGCACCCAGATGCA	165			
Db 1103	GACGCCCATATGTCTCTACAAATGATGTGGGTATAGAAAGGCCCAAGAACTTGT	1162			
Qy 166	GACACCTGTATGTAGAGAAAGATTGACGTGGGGAACTGAAA---CCACGACTGTGTC	222			
Db 1163	TCTGTTTGTGTCAAGCAATTTCTCTTGAGCCCTATGCAGAACCTTATGTTGTGTGA	1222			
Qy 223	AGCGGCCCTGTACAGTTCGTGCTCCCAAGAGGAACATGCAGAGCAGCGCTGATGCTCTG	282			
Db 1233	TGCACCTTGAACCAAGTTAACATGTAGAGGTATTAATCACTGCTGATGTGTGTGTT	1282			
Qy 283	TGCACCTCGAAACCCACAGAAAGATGAGAGGAGTCCAGATCCCAAGCCATAGCTTGTGTGCT	342			
Db 1283	TCTA 1286				
Qy 343	TCTA 346				

RESULT	9	SCYGL105w	1703 bp	DNA	PLN	11-AUG-1997
LOCUS						
DEFINITION		S.cerevisiae chromosome VII reading frame ORF YGL105w.				
ACCESSION		Z7627	YJ13135			
NID		g1322647				
KEYWORDS		baker's yeast.				
SOURCE		Saccharomyces cerevisiae				
ORGANISM		Eukaryote; mitochondrial eukaryotes; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
REFERENCE		1 (bases 1 to 1703)				
AUTHORS		Castagnoli, T., Paoluzzi, S. and Minenkova, O.				
JOURNAL		Unpublished				
REFERENCE		2 (bases 1 to 1703)				
AUTHORS		MIPS.				
TITLE		Direct Submission				
JOURNAL		Submitted (11-MAY-1996) Data collected by MIPS on behalf of the European yeast chromosome VII sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.emblnet.org				
FEATURES		Location/Qualifiers				
source		1..1703				
		/organism="Saccharomyces cerevisiae"				
		/db_xref="taxon:4932"				
		/chromosome="VII"				
		428..1558				
	gene	/gene="G4p1"				
		428..1558				
	CDS	/gene="G4p1"				
		/note="ORF YGL105w"				
		/codon_start=1				
		/db_xref="PID:e243983"				
		/db_xref="PID:g1322648"				
		/db_xref="SWISS-PROT:P4672"				
		/translation="MSDLYTKFESLIIKRVSTFKBSOAQAAOMESYLKSGIOPHL DQMLVBDNFTIVSTLTPYSTDVHVFVAPALIKDLYASCKDYKSTYTRHILRMIT DYNMLLEVSTSDKLEINHDLTDEHVELEKKKAPAGGAAPAAKADEDYSKAKKOD HPRKRPBEETIKLRLREKAKKAKKAAKNAKQOQKAKPEKTPSIDRVGF IQK AIRPMDADSLVSTIVDGEDGEPRTVSGLVKFLPDMADERYVAVVVCNLKIPVMRGI KSTPMVLGSGNDVDVEVEPEPKDSKAGKVFEEFGDEAPMKQIPLNPKRKIEHLQPHR				

RESULT	10				
LOCUS	SCAR1	2473 bp	DNA	UNA	14-OCT-1996
DEFINITION	S.cerevisiae ARc1 gene.				
ACCESSION	X95481				
NID	91620459				
KEYWORDS	ARc1 gene.				
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 2473)				
AUTHORS	Simos,G., Segref,A., Fasiolo,F., Hellmuth,K., Shervencko,A., Mann,R. and Hurt,E.C.				
TITLE	The yeast protein Arcp binds to tRNA and functions as a cofactor for the methionyl- and glutamyl-tRNA synthetases				
JOURNAL	EMBO J. 15 (19), 5437-5448 (1996)				
MEDLINE	97050848				
REFERENCE	2 (bases 1 to 2473)				
AUTHORS	Simos,G.				
TITLE	Direct Submission				
JOURNAL	Institue fuer Biochemie I, Im Neuenheimer Feld 328, 69120 Heidelberg, FRG				
COMMENT	Related sequence: U31348.				
FEATURES	Location/Qualifiers				
source	1..2473				
	/organism="unidentified"				
	/db_xref="taxon:32644"				
	1..2473				
	/organism="Saccharomyces"				
	/strain="J04.2XJR26.198"				
	/db_xref="taxon:4930"				
	554..1684				
	/gene="ARc1"				
	554..1684				
	/gene="ARc1"				
	/codon_start=1				
	/product="Arcp"				
	/db_xref="PID:e21389"				
	/db_xref="PID:g1620460"				
	/translation="MSDIYKFKFSLIISKYPVSTFKROSAQAAQESVLKSGIOPHLIDQNLVLRDNTFYVSTIYPTSDVHYFEVALPLKIDVASSDKYSTIYTRILNITDYMNQLVSVSTDLKLEIHNDLDPHEVIEIKKRPAGGADAALAKADEVSKKKKQDDHPRKDPDEETLKRLEAKAKAKAAKAAKQOQONKAPKRPALDVRGFIORAKIRPDDSYVSTIDVGDEGPRTVSGLVKHEPLDAMQERYVVVVCNKKPVMRGI				
gene					
CDS					

[illegible]

Db	10829	TCCACCTGGAACCGAGTAAATCAGAGTATTAATCCATCGGTATGTGAGT	10888
QY	283	TCCACCTGGAACCGAGTAAATCAGAGTATTAATCCATCGGTATGTGAGT	342
Db	10889	TCTA 10892	
QY	343	TCTA 346	
RESULT	12		
LOCUS	SCU31348	1745 bp	PLN 02-OCT-1995
DEFINITION	Saccharomyces cerevisiae G4p1 (g4p1) gene, complete cds.		
ACCESSION	U13348		
NTD	g1002711		
KEYWORDS	baker's yeast.		
SOURCE	Saccharomyces cerevisiae		
ORGANISM	Eukaryotes; mitochondrial eukaryotes; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.		
REFERENCE	1 (bases 1 to 1745)		
AUTHORS	Frantz,J.D. and Gilbert,W.		
TITLE	A novel yeast gene product, G4p1, with a specific affinity for quadruplex nucleic acids		
JOURNAL	J. Biol. Chem. 270 (35), 20692-20697 (1995)		
MEDLINE	95386521		
REFERENCE	2 (bases 1 to 1745)		
AUTHORS	Frantz,J.D.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-JUL-1995) J. Daniel Frantz, Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA		
FEATURES	Location/Qualifiers		
Source	1..1745		
gene	/organism="Saccharomyces cerevisiae"		
	/strain="S288c"		
	/db_xref="taxon:4932"		
	257..1384		
	/gene="g4p1"		
	257..1384		
	/gene="g4p1"		
	/note="G4 nucleic acid binding protein"		
	/codon_start=1		
	/product="G4p1"		
	/db_xref="pid:g1002712"		
	/translation="MSDLYVRFESLIISKYPSFTEKSOAOAWESVLKSGIOPHL DQMLVLRNMTFIVSTLPTSDVHVEVALPLIKDLVASSKDYKSTYTYRHLNWI DYONMLEVSTDKLEINHDLDPHEVIEKKRLRYALLMDOOGSDVSKKAKODH PRGPRDESLTKLKEEAKALLAKNAKNAOQOONKAEKPPASIDRVGFIOKA IKRHADPSLYVSTIDVGDDEGSPRVYCGVLVHFPLDMQERYVYVYCNLKPVMNRGIK STAVKDLGSDNDKEFEVPEPRDSKAGKVFEEGGDEAPMKOLNPKRKLIWEHLQPHFT TNDLEVIYFDEDEKRDHPVRKLTNAKESFKVASIANAOVR"		
BASE COUNT	582 a	326 c	345 g 492 t
ORIGIN			
Query Match	6.3%;	Score 32;	DB 18; Length 1745;
Best Local Similarity	61.8%;	Pred. No. 3.93e-05;	
Matches 123;	Conservative 0;	Mismatches 73;	Indels 3; Gaps 1;
Db	914	AAGCACCAGATGCGGACCTCATATGCTCTCAATGATGTCGGATGAAGAAGGC	973
QY	151	AAGCACCAGATGCGGACGAGCCTGTATATAGAGAAGATTGACGTGGGGAAGCTGAA	207
Db	974	CCAAAGACTGTTTGTCTGCTTGGTTCGCAAGCATTTTCTTTGGACGCTATGCAAGACGT	1033
QY	208	CCAGGACTGTGTGACGCGGCGCTGTACAGTTTCGTGCGCCCAAGAGCACTCAGACAGG	267
Db	1034	TATGTTGTTGTATGCAACTTGAACCACTTAACATAGAGAGGTATTAATCCACTGCT	1093
QY	268	CTGATGATGTGTGCTGTCAACTGTAACCCCAAGAAAGATGAGAGAGTGTGAGTCCCAAGGC	327
Db	1094	ATGATATGTGTGCTTCTA 1112	

RESULT	LOCUS	215 bp	DNA	PAT	30-OCT-1996
DEFINITION	Sequence 5 from patent US 5569830.				
ACCESSION	128278				
NID	91819054				
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 215)				
AUTHORS	Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.				
TITLE	Plant Inhibitors of fungal polygalacturonases and their use to control fungal disease				
JOURNAL	Patent: US 5569830-A 5 29-OCT-1996;				
FEATURES	Location/Qualifiers				
SOURCE	1..215				
BASE COUNT	15 a 8 c 25 g 26 t 141 others				
ORIGIN	/organism="unknown"				
Query Match	6.2% Score 31; DB 25; length 215;				
Best Local Similarity	14.3%; Pred. No. 1.59e-04;				
Matches	22; Conservative 64; Mismatches 67; Indels 1; Gaps 1;				
Db	1 MTNVTSSSVSRTASCNDKAKKDGNTTSSWTTDCNRTMGVCDTDTTRYVNDGHNK 60				
QY	CTGCTCTGTGTGACACGCTGTGTGAAGGCGTATGAAAGGCCAACACAGATGAGGAC 454				
Db	61 YSSANVTNGGNVGAAKHYTHYTVNSGDSKTYVDTSYASGTSSNGCTDGNRSGADSY 120				
QY	455 TCAACGCCAAGAAGAAAGCTTTCGAGAACTGCAGCGCTCACTCAAAATTCGTGAGAGCT 514				
Db	121 GSSKTAATSRNRTGKTANNVDSRNMGASVGS 154				
QY	515 GCATCGCACA-GTGGAGCAACCAACTTCATGA 547				
RESULT 14					
LOCUS	ECAE000389 11157 bp				
DEFINITION	Escherichia coli K-12 MG1655 section 279 of 400 of the complete genome.				
ACCESSION	AE000389 U00096				
NID	91789451				
KEYWORDS	Escherichia coli.				
SOURCE	Escherichia coli.				
ORGANISM	Escherichia coli.				
REFERENCE	Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.				
AUTHORS	1 (bases 1 to 11157)				
TITLE	Baltner,F.R., Plunkett III,G., Bloch,C.A., Perna,N.T., Burland,V., Rileyy,M., Collado-vides,J., Glasner,J.D., Rode,C.K., Maynew,G.F., Gregor,J., Davis,N.W., Kirpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.				
JOURNAL	The complete genome sequence of Escherichia coli K-12				
REFERENCE	Science 277 (5331), 1453-1474 (1997)				
TITLE	97426617				
JOURNAL	2 (bases 1 to 11157)				
AUTHORS	Baltner,F.R.				
REFERENCE	Direct Submission				
TITLE	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,				
JOURNAL	University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.				
REFERENCE	Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:				
TITLE	608-263-7459				
JOURNAL	3 (bases 1 to 11157)				
AUTHORS	Baltner,F.R.				
REFERENCE	Direct Submission				
TITLE	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,				
JOURNAL	University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.				
REFERENCE	Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:				
TITLE	608-263-7459				

COMMENT

The E. coli K-12 sequence and its annotations have been updated. All of the ambiguous residues in our original submission have been resolved, and mis-assemblies in two repetitive regions have been realigned. The annotations have been improved and updated as well. With this release we begin designating a version number for the annotated sequence, to assist in keeping track of corrections, updates, and other changes. This is version M52 (SEPT. 02, 1997). In addition, a revised notation has been instituted which assigns each gene (protein or RNA-encoding) a unique numeric identifier beginning with a lowercase 'b' (in the 'label' field); this will remain constant through further updates. gene identifications, etc. This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K-12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332. e-mail: mark@ember.gatech.edu

Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://gsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>).

location/Qualifiers

1. 11157

/organism="Escherichia coli"

/strain="K-12"

/sub_strain="MG1655"

/db_xref="taxon:562"

174. 797

/note="O207; This 207 aa ORF is 42 pct identical (7 gaps) to 52 residues of an approx. 200 aa protein SLYD_ECOLI SW: P30566"

/codon_start=1

/label-b3071

/db_xref="PID:g1789452"

/transl_table=1

/translation="MSHHNEGCCNHEGQPRHGGCCKEGSEHCHGHQHEHGCCGCGGRHGGRRGRREFGSGHGLVLDILSDPSHGVELIKAIENLTQGNTPSPVIVPTLIDLOESLITTRIEEGSGKQALITFGQAGMLEENEGYEMIERIKARCVGAALNQNPQKRALDNPKVALDPRVNSQSDSDQIKIIVADIRAFDITQLD"

complement(951. .2471)

/gene="air"

complement(951. .2471)

/gene="air"

/note="f506; 100 pct identical to AIR_ECOLI SW: P50466"

/codon_start=1

/label-b3072

/product="aerotaxis receptor protein"

/db_xref="PID:g1789453"

/transl_table=1

/translation="MSSHPVTOONTPLADDTILMSTDTLSQTYHANEFVQVSGYTLQELQGGPHNNVRHPDPKKAFAFDMFTLLKGEWMSGIYNNRRKNGHYIVRAAYAPMVRREKISGYMSIRTRATDEIDIAVEPIYKALMGRTSKRIHKGIVVRKGMGLKIPSLP LMRKAVGNTLMFTLLAAMLMFVAAPAVVYILCALVVLASACEMQIVRIENVAHQA LKAKATGERSNVEHLNRSDELGILTRAAGOLGMCRWLINDVKSIVSSVNGSETLAK GTYEDNTOOVNDVNOQVATNMOMASAKONCATISADSKITSTSNAAVOGCEAM TTYVKTNDIDTADSTORTGTTSLINDIAPONTILAAVAEARAGSGQVFAVAAEE VRLHASSANAANDIRKLIDASADKVSQGVQAHAAGRTMEDIVQYKNTQTLAQIS HSTLEQDQSSLLTRAVDELNLITQKNAELVEESAQVSAVYKRRASRLDEAAVYVLH"

2778. .4268

/gene="y9jg"

/gene="y9jg"

2778. .4268

/gene="y9jg"

/EC_number="2.6.1.13"

/note="O496; 100 pct identical to 429 amino acids of

OAT_ECOLI SW: P42588 but has 67 additional N-terminal residues"
 /codon_start=1
 /label=B3073
 /product="probable ornithine aminotransferase"
 /db_xref="pid:g1789454"
 /transl_table=11
 /translation="MITEVFRIPIPAANAQVQSIQYLRNHYRREPETHLNLNPPSSASALACSNHALILKRTILDEEMKALREVIETKEHVAFGFLEIKRSVYAGDYGAVEMQAGSLNLTVDVTOGQEFIDCLGSGFLEFNQHRPNVVASVQNOQLAHPHSGELDVPRAMILAKTLAIPGKILKSYFCNCSGTSVEALAKLAKASRGKPTFYATSGAIFRGLSIALSTASTFRKPFEMLLPGFRHVPGENIPAMRTALNECKTIGDVAATLEFRIOGEGVILPPEGYIATARKCDERGAALMLDEVOTGNGPKFACGHEWVODIICLAKAGCGMIPGATIAATREKVSULFDPNPELHTTGGSNPLACAAALATIVLVEQLNPAAQKQGMULDGERQULAREVPDLVQERKGMALAEFDNEIGINFASEMFRQVRNVAQGLNNAKTIRIPPLTLTIEOCELVIKARKALAMRVSEAA"
 complement(4310..4642)
 /gene="y9jH"
 complement(4310..4642)
 /gene="y9jH"
 complement(4310..4642)
 /gene="y9jH"
 /note="f110: 100 pct identical to YGJH_ECOLI SW: P42589"
 /codon_start=1
 /label=B3074
 /product="hypothetical 12.3 kb protein in llex-ebgr intergenic region"
 /db_xref="pid:g1789455"
 /transl_table=11
 /translation="METVAYADFARLENRVKRIYEVKRHENADKLIIYVDVQGTQLQITVTSIAPVYSEBELMKTIVVVLGNLQKAKMGETSECMILCAETDDGSESVLLPERMMPEAGRVVV"
 4861..5844
 /gene="ebgr"
 4861..5844
 /gene="ebgr"
 /function="ebg repressor"
 /note="c327: 100 pct to EBGR_ECOLI SW: P06846; CG Site No. 829"
 /codon_start=1
 /label=B3075
 /db_xref="pid:g1789456"
 /transl_table=11
 /translation="MATLKDAIEAGVSLATVSRANDPPTLVNKEETKHLILIAERKLETKSSARKIOTGAVNOHHILAIYSOOEELNDPYLARIHGIEOTCEADLIGLIELNRCEYHSGLPDIKVTGILIVGKPTPALRAASLTDNICTFIDFHEGSGTIDVDILNKRISKEILIDFINGVNRIGETIGEDGPRADIREVAFAEYGRKQVREEDIMWRGFESSSSGYELAKOMLAREDPALFYASDSIAIGVLRAIHERGLNIPDILSIYNDPIPARFTFPLSTVRIHSEMKSGQVNLVYEKARDGRALPLIVFSPSKLKLKRTGR"
 5992..9120
 /gene="ebga"
 5992..9120
 /gene="ebga"
 /note="o1042: This 1042 aa ORF is 99 pct identical (1 gaps) to 1032 residues of an approx. 1032 aa protein BGA2_ECOLI SW: P06864; CG Site No. 830"
 /codon_start=1
 /label=B3076
 /product="phospho-beta-D-galactosidase: alpha-subunit"
 /db_xref="pid:g1789457"
 /transl_table=11
 /translation="MNGSDFSLPGGPMNRWENIQLTHENRLAPAFESYVQAAQRTFARHSSFLPRLSGOMNHFHDPRLQVYEATSELMADWGHTTPRAMQMGHKLQYIDEGEPFLDIDVDFVSDNPTGAYQRIIFLISGOWMOKQYLLIFDGGEITYEIVYAGVGOYDGSKSRPLAEFEDISAMWKTGDNILCVVNEAMSTVEDDMMMSAIFEDYVALVKRHLIINDFVTFDPEADATLSCSEVLENLASPVYVLTLEYLFDERVYSALIVHAIKRLISATPAFVLEQPOOMSAESPYLHLYMVLKRNANGVLEVPDORGFSDIVDGLTWINNRYMLAGVRNDRDKRGAVGCDRERKQLQMKOHNINSVYTAHPMDPREFELCDIYGLFVAEIDVBSHGFAVNGDISKITLDQMKREYVERIVRIIAQNNHNSPILIIWSGNEGSGYGCNIRAMYHAKARALDRLRLVYEDRQAEVYDIISTATYTPVLMNENFGRPHKPRILICEYHANGMGNGGLTEYOPNGKCDJCGHAYWEMCHQIADQDNDFGNYVAFKGGVGLGPPNNVANGCLDGLIFYSOOTPEPGKQVYAPAKHARLFTGSGELKVENKIMETITLDYGLHAAYRAGSELAKIQIKLRVYANSRBPLOITLPQADAREALVNTYIKDSTRSEAGHPLATIQYFLKRETAQPVAFNARNRPLULEDRDLSCTIVIKQYNIATFYSKSGSPITSMQVNGESLLIREPILNFKMDINHAKOEYEGMLQPHNLDI

Note: remainder of annotations omitted.

Query Match	6.08;	Score 30;	DB 19;	Length 11157;
-------------	-------	-----------	--------	---------------

Matches 76; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Db 4386 GCACACAACAGCATGCAITTCGACGTTTCACCACGCAITTTTCGCTTTTGCAGATTGCAT 4445

CP 341 GCACACAGAAGCATGCCCTGGGACTGACTCCTCTCATCTTCTGGGTTTCAGGTGCAC 282

Db 4446 AAGACCACCACCGTTTCCCATCAGCTCTTCTTGGCTGTAGTACGGCACCAGGCTGTC 4505

CP 281 AGCACCTACCA GCCCTGTCCTGCAGTTCCTCCTGGGACGAACTGTACCAGGCCGCTC 222

Db 4506 AC 4507

CP	221	AC	220
----	-----	----	-----

gene

genes

CDS

MOEHLPBEAAGSNGSEVLLISPIVPIAPVPFSDGMSQATYIMRIADAGOVNVALSGEERFG
DYPIHPIECIGMTKINIEHYNOQVAYVRRGEGENTADSCQATYIDIRACVADMPENFYEG
PONGNRQHVWAKMLNIRKSGEYVYVQKIRFNSAMHTQIENIAHQCHQASDDSDI
TILNIDHOLGSGWSEVLSDMRWVRFDSYFTGLLPVSGGENTAGSLATSYEFEGAGQ
FEFTNLSEHNQK"
9117. 9566
/gene="ebgc"
9117. 9566
/gene="ebgc"
/note="O149; CG Site No. 18412; difference near end makes
this 25 amino acids shorter than previously published
sequences; 100 pct identical amino acid sequence and equal
length to EBG_ECOLI SM: P24225"
/codon.start=1
/label-b3077
/product="Phospho-beta-D-galactosidase, beta-subunit"
/db_xref="PIR:g1789458"
/translation="MRTIINDLEQFOIYASGKKMORCEALINIDIQIOPVAHSIGSS
LTVREYDSDALDFGHRREYFVHYVYLOGOKIKAPENILOVVEYRBDTDEYLYK
SGGEYVEYHSGQIYCIDIHAIYFICGNNAVKRYVLTQVITLEDGYFHNK"
9629. 11062

RESULT	15		
LOCUS	ECOUR67	372438 bp	DNA
DEFINITION	Escherichia coli K-12 chromosomal region from 67.4 to 76.0 minutes		

NID 9606010

SOURCE	Escherichia coli.
1. Milk	100
2. Meat	100
3. Eggs	100
4. Vegetables	100
5. Fruit	100
6. Water	100
7. Air	100
8. Soil	100
9. Sewage	100
10. Human feces	100
11. Animal feces	100
12. Urine	100
13. Saliva	100
14. Sweat	100
15. Tears	100
16. Blood	100
17. Pus	100
18. Discharge	100
19. Urine	100
20. Feces	100
21. Sputum	100
22. Vaginal discharge	100
23. Menstrual discharge	100
24. Semen	100
25. Urine	100
26. Feces	100
27. Sputum	100
28. Vaginal discharge	100
29. Menstrual discharge	100
30. Semen	100
31. Urine	100
32. Feces	100
33. Sputum	100
34. Vaginal discharge	100
35. Menstrual discharge	100
36. Semen	100
37. Urine	100
38. Feces	100
39. Sputum	100
40. Vaginal discharge	100
41. Menstrual discharge	100
42. Semen	100
43. Urine	100
44. Feces	100
45. Sputum	100
46. Vaginal discharge	100
47. Menstrual discharge	100
48. Semen	100
49. Urine	100
50. Feces	100
51. Sputum	100
52. Vaginal discharge	100
53. Menstrual discharge	100
54. Semen	100
55. Urine	100
56. Feces	100
57. Sputum	100
58. Vaginal discharge	100
59. Menstrual discharge	100
60. Semen	100
61. Urine	100
62. Feces	100
63. Sputum	100
64. Vaginal discharge	100
65. Menstrual discharge	100
66. Semen	100
67. Urine	100
68. Feces	100
69. Sputum	100
70. Vaginal discharge	100
71. Menstrual discharge	100
72. Semen	100
73. Urine	100
74. Feces	100
75. Sputum	100
76. Vaginal discharge	100
77. Menstrual discharge	100
78. Semen	100
79. Urine	100
80. Feces	100
81. Sputum	100
82. Vaginal discharge	100
83. Menstrual discharge	100
84. Semen	100
85. Urine	100
86. Feces	100
87. Sputum	100
88. Vaginal discharge	100
89. Menstrual discharge	100
90. Semen	100
91. Urine	100
92. Feces	100
93. Sputum	100
94. Vaginal discharge	100
95. Menstrual discharge	100
96. Semen	100
97. Urine	100
98. Feces	100
99. Sputum	100
100. Vaginal discharge	100

Eubacteria; Prote

ESCHERICH, J.
1 (bases 1 to 37)
REFERENCE

AUTHORS PLUNKETT, G.
TITLE none

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3)

Plunkett, G.
Director, Subsector

JOURNAL
Submitted (22-DEC-2019)

COMMENT This sequence was

Wisconsin-Madison

from Escherichia

start of the enti

Location/qualifiers

```

1: 0/0/0
/organism="Escherichia coli"

```


	ORGANISM	Homo sapiens Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homidae; Homo.
	REFERENCE	1 (bases 1 to 525) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Matra,M., Pearson,J., Rifkin,I., Rohlfing,T., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R. Unpublished (1995)
	TITLE	Contact: Wilson RK
	JOURNAL	Mashu-Merck EST Project
	COMMENT	Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LMLT ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -28M13 rev7 from Amershamb High quality sequence stop: 330. Location/Qualifiers .size 1..525 /organism="Homo sapiens" /note="Organ: colon; Vector: pBluescript SK-; Site.1: EcORI; Site.2: XhoI; Cloned unidirectionally. Primer: Oligo dt. T-84 colonic epithelial cell line". Average Insert size: 1.0 kb; Uni-ZAP XR Vector; ~5 adaptor sequence: 5' GAATTCGACAGc 3'-3' adaptor sequence: 5' CTCGAGTTTTTTTGTGGTTTTTTT 3'. " /db_xref="taxon:9606" /_clone="511328" /_clone.lib="-Stratagene colon (#937204)" /_lab_host="SOLR cells (kanamycin resistant)" <1..>525 BASE COUNT mRNA ORIGIN 144 a 125 c 149 g 105 t 2 others
Db	Query Match	60.8%; Score 363; DB 15; Length 525;
	Best Local Similarity 96.5%;	Pred. No. 0.00e+00;
Matches	412; Conservative	0; Mismatches 8; Indels 7; Gaps 7
Db	62 TACCCTGCCTGAAAAAAACTGCCAGCAGCTCCATTACCAGATCCCTCAAGCAGAACC	121
OY	1 TACCCCCTGCCTGAAAAAACTGCCAGCAGCTCCATTACCAGATCCCTCAAGCAGAAGCC	60
Db	122 AATGNCAAAACCCTATGCCAATAATTGACAACCAAAGAGCATATCCCATCCCCGCTGAT	181
	61 AATGGCCAAAAGCCTGCCMAGAATTCAAGAACCAAGAGGATCATCCATCCCGCTGGAT	120
Db	182 ATCCGTGTGGGAAAAATCATCATCTGTGGAGAGAACCAGATGCAGACAGCCTGTATGTA	241
OY	121 ATTCCGTGTGGGAAAAATCATCATCTGTGGAGAGAACCAGATGCAGACAGCCTGTATGTA	180
Db	242 GAGAAATTGACGTGGGGGAAAGCTGAACACAGACTGTGTGTAAGCGGCTGTACAGTT	301
	181 GAGAAAGATTGACGTGGGGGAAAGCTGAACACAGACTGTGTGTAAGCGGCTGTACAGTT	239
Db	302 CGTGCCCAAGAGAAATTCAGCAGGACAGCGCTGTGTGTGTGTGTGCAACCTGAACCCA	361
OY	240 C GTGCCCAAGAGAAATTCAGCAGGACAGCGCTGTGTGTGTGTGTGTGCAACCTGAACCCA	299
Db	362 GAAGATGAGAGGAGTAGTCACCAAGGCATGCTTGTGTGTGTGCTTCATAGAAAGGATA	421
OY	300 GAAATGAGAGG - AGTCAGATGCCAAGGCATGCTTGTGTGTGTGCTTCATAGAA - GGAGATA	357
Db	422 AACGCCAGTTGAACCTCTGCAACCTCCGGCAGAGCTCTGTCTCCCTGGTTGACAGCT	481
OY	358 AACGCCAGGTTGAACCTCT - GCACCTCCGGCA - GGCTGTGCTCC - TGCT - GAGCACGT	413
Db	482 TTGTTTG	AAG

OY	414	GTTTGTG	420
RESULT	4	R18984	407 bp mRNA EST 14-APR-1995
LOCUS		yq25d10.r1 Homo sapiens cDNA clone 33689 5' similar to SP:SYM_THERRH	
DEFINITION		P23395 METHIONYL-TRNA SYNTHETASE ;.	
ACCESSION		R18984	
NID		g772594	
KEYWORDS		EST.	
SOURCE		human cDNA--33689 library-Souares infant brain INIB vector-Lafmid BA host-PHI10B (ampicillin resistant) primer-M13RP1 RstEcl-Not I RstEcl-Hind III Whole brain from a 73 days post natal female. 1st strand cDNA was primed with a Not I - Oligo(dT) primer [5'] AACTGGAAATATTCGGCGCCAGGAATTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonalio.	
ORGANISM		Homo sapiens Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	
REFERENCE		1 (bases 1 to 407)	
AUTHORS		Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaaski,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.	
TITLE		The WashU-Merck EST Project	
JOURNAL		Unpublished (1995)	
COMMENT		GDB: G00-406-036 Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu High quality sequence stops: 294 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Location/Qualifiers 1..407 /organism="Homo sapiens" /clone="33689"	
FEATURES		source	
BASE COUNT		103 a 87 c 121 g 95 t 1 others	
ORIGIN			
Query Match		48.4%; Score 289; DB 10; Length 407;	
Best Local Similarity		94.5%; Pred. No. 0.00e+00;	
Matches 364; Conservative		0; Mismatches 10; Indels 11; Gaps 11;	
Db	1	AAGATTGACGTGGGGGAAGCTGAACACAGCATGTGTGATGAGCGGCTGTGACAGTTCGTG 60	
Oy	184	AAGAATTGACGTGGGGGAAGCTGAACACAGCATGTGTGATGAGCGGCTGTGACAGTTCGTG 243	
Db	61	CCCAAGAGGAACTGCAGAGACAGGCTGTGTCTGTCTGTCACAACTTGAAAACCCAGAG 120	
Oy	244	CCCAGAGGAGAACTGCAGAGACAGGCTGTGTCTGTCTGTCACAACTTGAAAACCCAGAG 303	
Db	121	ATGAGAGAGTCGATGCCAAGGCAATGCTTGTGTGCTTCTATAGAAAGGATAAACCGC 180	
Oy	304	ATGAGAGAGTCGATGCCAAGGCAATGCTTGTGTGCTTCTATAGAAAGGATAAACCGC 363	
Db	181	CAGTTGANAACCTTGAGACCTTCGCGAGAGCTTGTGCTTCTCCGTGGGTGAGCACAGTGTGTG 240	
Oy	364	CAGGTGAAACCTTGAGACCTTCGCGAGAGCTTGTGCTTCTCCGTGGGTGAGCACAGTGTGTG 420	
Db	241	AAGGCTTATGAAAAAGGCCAACCAAGATGAGAGCTCAAGCCCAAGAAAGAAAGTCTTTTGA 300	

```

OY 421 AAGGGCTATGAAAAAGGCCAACAGATGAGAGCTCAAGCCCAAGAAAGTCTT-CGA 479
Db 301 GAAGTTGACGGCTGACTTCAAAATTTCTTGAGAGTGATCGCAAGTTGGGAACA 360
OY 480 GAAGTT-GACGGCT-GACTT-CAAAATTTCT-GAGGAGTGAT-CGCACAGTGG-AAGCA 533
Db 361 AACCACTTTCATGACCAAGTTGGG 385
OY 534 AACCACTTCAT-GACCAAGCTGGG 557

RESULT 5
LOCUS AA102052 459 bp mRNA 28-OCT-1996
DEFINITION 2178g12.r1 Stratagene colon (#937204) Homo sapiens CDNA clone
510790 5' similar to WP:F58B3.5 CE06007 METHIONYL-TRNA SYNTHETASE
;
ACCESSION AA102052
NID g1645892
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 459)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevasakis,E.,
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
Waterson-Merc EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT
Contact: Wilson RK
WashU-Merc EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seg primer: -28m13 rev2 from Amersham
High quality sequence stop: 418.
Location/Qualifiers
1. 459
/organism="Homo sapiens"
/note="Organ: colon; Vector: pBluescript SK-; Site.1:
EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer:
Oligo dt: T-84 colonic epithelial cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGCG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3'"
/db_xref="taxon:9606"
/clone="510790"
/clone_lib="Stratagene colon (#937204)"
/lab_host="SOLR cells (kanamycin resistant)"
<1..>459

BASE COUNT 114 a 118 c 117 g 107 t 3 others
ORIGIN
mRNA
Query Match 46.6%; Score 278; DB 15; Length 459;
Best Local Similarity 95.6%; Pred. No. 0.00e+00;
Matches 368; Conservative 0; Mismatches 2; Indels 15; Gaps 15;

Db 1 AAGCGGCTGTACAGTTGCGGCCCAAGAGAGTCAAGCGCTGTAGTGTGCTG 60
OY 223 AGCGGCTGTACAGTTGCGGCCCAAGAGAGTCAAGCGCTGTAGTGTGCTG 282
Db 61 TCGAACCTGAACCCCAAGAGATGAGAGAGTCAAGCGCTGTAGTGTGCTG 120
OY 283 TCGAACCTGAACCCCAAGAGATGAGAGAGTCAAGCGCTGTAGTGTGCTG 342
Db 121 TCTATGAGAGGATTAACCGCCAG-TTGAACCTGTGACCTCCGCGAG-CTCTGCTCT 178

```

```

OY 343 TCTATGAGAGGATTAACCGCCAGGTGAACCTCTGGACCCCTCCGGCAGGCTGCTGCT 402
Db 179 GGTGAGACCGCTTTTGTGAAGGCTATGAAGGGCCAAACAGATGAGAGTCAAGCC 238
OY 403 GGTGAGAC-GTGTGTTGTGAAGGCTATGAAGGGCCAAACAGATGAGAGTCAAGCC 461
Db 239 CAAGAGAAGAGTCTTGAGAAAGTCAAGCTGACTTCAAAATTTCTTGAGAGAGTGCAT 298
OY 462 CAGAAGAAGAGTCTTGAGAAAGTCAAGCTGACTTCAAAATTTCTTGAGAGAGTGCAT 518
Db 299 CACCAAGTTGGAACCAACCAACTTCATGACCAAGCTTGGCTTCATTTCCCTGTTA 358
OY 519 CGCACA-GT-GCAAGCAAAACCAACTT-CAT-GACCAAGCT-GGGCTCAATTTCC-TGT-A 571
Db 359 AATCGCTGAAGGGGGGACACTT 383
OY 572 AATCGCT-GAAG-GGGGGACACTT 594

RESULT 6
LOCUS AA221081 517 bp mRNA 12-FEB-1997
DEFINITION mw65f03.r1 Soares mouse 3NME12.5 Mus musculus CDNA clone 659933 5'
similar to TR:G1184699 G1184699 TYROSYL-TRNA SYNTHETASE. ;
ACCESSION AA221081
NID g1840276
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 517)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Matlin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)

TITLE
JOURNAL
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouse@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seg primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 493.
Location/Qualifiers
1. 517
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pT73D-Pac (pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand CDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCACATCATGTGAGAGAGCGGCGCCGCTTATTTTTTTTTTTT 3'],
on total mouse RNA (provided by Minoru Ko, Wayne State
Univ.); double-stranded CDNA was ligated to Eco RI
adaptors (pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:10090"
/clone="659933"
/clone_lib="Soares mouse 3NME12 5"
/sex="unknown"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
<1..>517

mRNA

```



```

source
1..243
/organism="Homo sapiens"
/note="vector: lambda ZAP Express; Site_1: EcoRI; Site_2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dt adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."
/db_xref="taxon:9606"
/clone="U1914"
/clone_lib="Fetal heart, lambda ZAP Express"
/lab_host="E. coli XL1-Blue"
<1..>243
BASE COUNT      57 a    58 C    77 G    51 t

Query Match      36.7%; Score 219; DB 14; Length 243;
Best Local Similarity 96.7%; Pred. No. 0.00e+00;
Matches 237; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

Db 1 CCCAGATGCACACAGCCCTGATGTAGAGAAGATTGACGTGGGGAMCTGAACCGAGC 60
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 156 CCCAGATGCACACAGCCCTGATGTAGAGAAGATTGACGTGGGGAMCTGAACCGAGC 215
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 61 TGTGGGACGGCGCTGTACAGTTCGTGCCCAAGGAGAACTCGAGCACAGCGTGAGT 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 216 TTGTGGAGACGGCGCTGTACAGTTCGTGCCCAAGGAGAACTCGAGCACAGCGTGAGT 275
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 121 TTGTGCTTTCAACCTGAACCACCCAGAAATGAGAGAGTAGTC-AAGCATGCTTCT 179
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 276 GGTCGCTGTCAACCTGAACCACCCAGAAATGAGAGAGTAGTCAGATCCCAAGCATGCTTCT 335
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 180 GTGTCTTCTATAGAAGGATAA-CGCCAGGTTGAACCTCTTGCCCCCTCGGACAGTTTC 238
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 336 GTGTCTTCTATAGAAGGATAAACCGCCAGGTTGAACCTCTGACCCCTCGGACAGGCTC 395
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 239 TGCTC 243
|||||
QY 396 TGCTC 400

RESULT          9
LOCUS           AA003398         423 bp      mRNA      EST       19-JUL-1996
DEFINITION     mg57b09.r1 Soares mouse embryo NBME13.5 I4.5 Mus musculus CDNA
                clone 437081 5' similar to PIR.A55053 A55053
                endothelialmonocyte-activating protein II precursor - mouse ; .
ACCESSION      AA003398
NID            g1446863
KEYWORDS
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Mitochondrial eukaryotes; Metazoa; Chordata;
               Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
               Murinae; Mus
REFERENCE      1 (bases 1 to 423)
AUTHORS      Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
              Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
              Schellenberg,K., Steptoe,M., Tan,F., Underwood,R., Moore,B.,
              Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
              Waterston,R.
TITLE         The WashU-HM Mouse EST Project
JOURNAL       Unpublished (1996)
COMMENT
Contact: Maria M/Mouse EST Project
WashU-HM Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LMU ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:262417

```

Seq primer: ETPRimer
High quality sequence stop: 343.
Location/Qualifiers
1. 423
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - Oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCGGAATTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dp and 2 14.5dp embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 1] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3D vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaudo."

/db_xref="taxon:10090"
/clone="437081"
/clone_1lb="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dp total fetus"
/lab_note="DH10B"
<1. >423

BASE COUNT 98 a 115 c 120 g 90 t

ORIGIN

Query Match 36.0%; Score 215; DB 14; Length 423;
Best Local Similarity 90.3%; Pzed. No. 0.00e+00;
Matches 241; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Db 157 GAAGTGCAGAGAGAGGCTGTGTGTGTGTGTGCAATCTGAATCCAGAGATGAGAGC 216
|||||
Qy 253 GAAGTGCAGAGAGAGGCTGTGTGTGTGTGTGCAATCTGAATCCAGAGATGAGAGC 312
|||||

Db 217 GTGAGACTGGCAGAGGCTACTGTGTCTTCTGTAGAAGGGGTGAGCCCGCAGTTGAA 276
|||||
Qy 313 GTGAGATCCCAAGGCATCTTCTGTGTCTTCTGTAGAAGGGAATAACCCGAGTTGAA 372
|||||

Db 277 CCTGTGAGACCCCTCGTGGGTGCTCTCTCGTGTGAGACGATGCTGACAGGGCTATGAG 336
|||||
Qy 373 CCTGTGAGACCCCTCGGAGGCTCTCTCTCTGTGTGAGACGATGCTGAGAGGCTATGAA 432
|||||

Db 337 AAGGGCCAGCCAGATGAGAGAGCTCAAGCCCAAGAAAGATCTTTGAGAAAGCTCAGGCC 396
|||||
Qy 433 AAGGGCCCAACCGATGAGAGAGCTCAAGCCCAAGAAAGATCTTGAAGATTCAGAGCT 492
|||||

Db 397 GACTTTAAATTTCTGAGAGATGCATC 423
|||||
Qy 493 GACTTCAAAATTTCTGAGAGATGCATC 519
|||||

RESULT 10

LOCUS H13315 437 bp mRNA EST 27-JUN-1995
DEFINITION y112b07.r1 Homo sapiens cDNA clone 43366 5' similar to
SP:SYSC_YEAST P36421 TYROSYL-TRNA SYNTHETASE, CYTOSOLPLASMIC ; .
H13315
ACCESSION
NID 9878135
KEYWORDS
SOURCE EST.
human clone-43366 library-Soares infant brain 1N1B vector-lafmid BA
host-DH10B (ampicillin resistant) primer-M13P1 Rstfel-Not I
strand-2Hind III Whole brain from a 73 days post natal female. 1st
strand cDNA was primed with a Not I - Oligo(dT) primer [5'
AAGTGAAGATTCGGCGCCGAGCAATTTTTTTTTTTTTTTT 3']; double-stranded
cDNA was ligated to Hind III adaptors (Pharmacia), digested with
Not I and directionally cloned into the Not I and Hind III sites of
the lafmid BA vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.Fatima
Bonaudo.
Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;

REFERENCE
AUTHORS

Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.

TITLE
JOURNAL
COMMENT

The WashU-Merck EST Project
Unpublished (1995)

GDB: G00-415-907
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watsn.wustl.edu
High quality sequence stops: 398
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Location/Qualifiers

FEATURES

1..437
/organism="Homo sapiens"
/clone="43366"

BASE COUNT
ORIGIN

126 a 97 c 116 g 95 t 3 others

Query Match

Best Local Similarity 98.1%; Pred. No. 0.00e+00;
Matches 212; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

Db

220 TACCCCTGCCCTGAAAAAAGTGGCCAGCGCTGCTACCCAGATCCCTCAAGAAGACC 279

Qy

1 TACCCCTGCCCTGAAAAAAGTGGCCAGCGCTGCTACCCAGATCCCTCAAGAAGACC 60

Db

280 AATGGCCAAAGCGCTGCGCAAGATTCAGAACCCAGAGAGAGTATCCCATCCCGGCTGGA 339

Qy

61 AATGGCCAAAGCGCTGCGCAAGATTCAGAACCCAGAGAGAGTATCCCATCCCGGCTGGA 119

Db

340 TATCCCTGTGGGAGANATCATCTGTGAGAGACACCCAGATGAGAGAGAGCTGATG 399

Qy

120 TATCCCTGTGGGAGANATCATCTGTGAGAGAGACACCCAGATG-CAGACAGCTGTATG 178

Db

400 TAGAGAAGATTGACGTGGGGAAGCTGAACACGGA 435

Qy

179 TAGAGAAGATTGACGTGGGGAAGCTGAACACGGA 214

RESULT 11

LOCUS AA327316 248 bp mRNA EST 20-APR-1997
DEFINITION EST10628 Colon 1 Homo sapiens cDNA 5' end.
ACCESSION AA327316
NID 91979561
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae;

REFERENCE
AUTHORS

1 (bases 1 to 248)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man,Wai,C.,
Clifton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Heblum,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-T., Marmaro,S.M., Merrick,J.M.,
Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,

TITLE
JOURNAL
MEDLINE
COMMENT

Small,R.V., Springs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bedarick,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.R.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl.), 3-174 (1995)

Other ESTs: THC181443
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES

1..248
/organism="Homo sapiens"
/note="Organ: colon; Vector: pBluescript SK-; Site-1:
ECORI; Site-2: XhoI"
/db_xref="ATCC (Inhost):127993"
/db_xref="ATCC (Inhost):127993"
/db_xref="taxon:9606"
/clone_1fb="Colon 1"
/dev_stage="adult"

BASE COUNT
ORIGIN

67 a 63 c 60 g 55 t 3 others

Query Match

Best Local Similarity 97.5%; Pred. No. 0.00e+00;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db

1 CTCCTGTGAGACAGCTGTTTGAAGGCTATGAAGGCGCCACACGATGAGAGCTCA 60

Qy

398 CTCCTGTGAGACAGCTGTTTGAAGGCTATGAAGGCGCCACACGATGAGAGCTCA 457

Db

61 AGCCCAAGAGAAAGTNTTCAGAGAGTTCAGAGCTGACTTCAAAATTTCTNAGAGTGA 120

Qy

458 AGCCCAAGAGAAAGTNTTCAGAGAGTTCAGAGCTGACTTCAAAATTTCTNAGAGTGA 517

Db

121 TCGCTGAGTGAAGCAACCACTTCATGACCAAGCTGGCTCCATTCTCTGAATGCG 180

Qy

518 TCGCTGAGTGAAGCAACCACTTCATGACCAAGCTGGCTCCATTCTCTGAATGCG 577

Db

181 TGAAGGGGGAACATTAGC 200

Qy

578 TGAAGGGGGAACATTAGC 597

RESULT 12

LOCUS AA230715 503 bp mRNA EST 26-FEB-1997
DEFINITION NM04908.1 Soares mouse 3NE12 5 Mus musculus cDNA clone 669758 5'
similar to TR:G1184699 G1184699 TYROSYL-TRNA SYNTHETASE. ;
ACCESSION AA230715
NID 91853030
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE
AUTHORS

1 (bases 1 to 503)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)

CONTACT: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:409462
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 434.
Location/Qualifiers
1..503
/organism="Mus musculus"
/strain="C57BL/6J"
/note="vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCGCTTATTTTCTTTT 3'],
on total mouse RNA [provided by Minoru Ko, Wayne State
Univ.]; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:10090"
/clone="669758"
/clone_lib="Soares mouse 3MME12 5"
/sex="unknown"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
1..>503

BASE COUNT 139 a 121 c 141 g 102 t
ORIGIN
Query Match 27.6%; Score 165; DB 17; Length 503;
Best Local Similarity 89.4%; Pred. No. 2,646-283;
Matches 193; Conservative 0; Mismatches 22; Indels 1; Gaps 1;

Db 288 TATCCAGCCCTGGAAGAGCTGGCCAGTGCCTACCACTCTTAAAGCAGAACCC 347
|||||
QY 1 TACCCCTGCTGTAAGAAATGCGCAGCGCTGCTTACCACTCTTAAAGCAGAACCC 60
Db 348 TCTGCAAGAGCCCTGCAAGACTGCAAGCAGAGAGAGCATCCATCCGGCTGGA 407
|||||
QY 61 AATGGCCAAAGGCC-TGCCAAGAAATTCGAAGCAGAGAGGTCAATCCCGCTGGA 119
Db 408 TATCCGTGTGGGCAAAATCTCAGCGTGGAGAGCACCAGATGCATAGCTTTATGT 467
|||||
QY 120 TATCCGTGTGGGCAAAATCATCATCTGTGGAGAGACCAATGATGAGACCTGTATGT 179
Db 468 GGAGAAGATTGATGTGGGGAGAGCTGAACCAAGAC 503
|||||
QY 180 AGAGAAGATTGATGTGGGGAGAGCTGAACCAAGAC 215

RESULT 13
LOCUS H22113 209 bp mRNA EST 06-JUL-1995
DEFINITION Y134906.r1 Homo sapiens cDNA clone 160186 5'.
ACCESSION H22113
NID 9890808
KEYWORDS EST.
SOURCE human clone-160186 library-Soares breast 3NBHST vector-pT7T3D
(Pharmacia) with a modified polylinker host-DH10B (ampicillin
resistant) primer-M13RP1 Rsite1-Not I Rsite2-Eco RI Adult human.
1st strand cDNA was primed with a Not I - oligo(dT) primer [5'

RGTTACCAATCTGAAGTGGAGCGCGCCGCTTATTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
digested with Not I and cloned into the Not I and Eco RI sites of a
modified pT7T3 vector (Pharmacia). Library went through one round
of normalization to a Cot = 20. Library constructed by Bento Soares
and M. Fatima Bonaldo.
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 209)
REFERENCE
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

CONTACT: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
High quality sequence stops: 162
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Location/Qualifiers
1..209
/organism="Homo sapiens"
/clone="160186"

BASE COUNT 60 a 56 c 47 g 42 t 4 others
ORIGIN
Query Match 27.3%; Score 163; DB 11; Length 209;
Best Local Similarity 97.7%; Pred. No. 5,016-219;
Matches 169; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Db 1 GCTATGAAAGGGCCACAGATGAGAGCTCAAGCCCAAGAGAGTCTTCAGAGT 60
|||||
QY 425 CCTATGAAAGGGCCACAGATGAGAGCTCAAGCCCAAGAGAGTCTTCAGAGT 484
QY 485 TGACGCTGACTTCAAAATTTCTGAGAGTGCATCGC-CAGTNGANCACCAACTTCA 119
Db 120 TGACCAAGCTGGGCTCATTTCTGTAAATCGCTGAAAGGGGGAGACATTAAGC 172
|||||
QY 545 TGACCAAGCTGGGCTCATTTCTGTAAATCGCTGAAAGGGGGAGACATTAAGC 597

RESULT 14
LOCUS AA049626 500 bp mRNA EST 09-SEP-1996
DEFINITION m36006.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
ACCESSION AA049626
NID 91529297
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Eukaryota; Eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
1 (bases 1 to 500)
REFERENCE
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

WITNESS
(TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSPrch_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sat Apr 18 02:04:34 1998; Maspar time 11.13 Seconds

Tabular output not generated. 646.917 Million cell updates/sec

Title: >US-08-483-534A-2
Description: (1-168) from US08483534A.pep
Perfect Score: 1198
Sequence: 1 EEVIFPSRLDIRVGKIIIVK.....NFMTKIGSISCKSLKGNIS 168

Scoring table:
PAM 150
Gap 11

Searched: 195121 segs, 42852602 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: plr55
1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 36.208; Variance 118.180; scale 0.306

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	596	49.7	310	2	A55053 This is a DE line.	1.52e-58
2	588	49.1	310	2	B55053 This is a DE line.	1.54e-57
3	462	38.6	376	2	S64113 This is a DE line.	8.01e-42
4	308	25.7	110	2	G65095 This is a DE line.	4.30e-23
5	266	22.2	616	1	SYTWT This is a DE line.	3.87e-18
6	263	22.0	658	2	D69431 This is a DE line.	8.67e-18
7	222	18.5	651	2	G69177 This is a DE line.	4.50e-13
8	211	17.6	651	2	P64457 This is a DE line.	8.06e-12
9	200	16.7	682	2	I64113 This is a DE line.	1.38e-10
10	188	15.7	649	2	S16682 This is a DE line.	2.95e-09
11	181	15.1	664	2	S66067 This is a DE line.	1.73e-08
12	181	15.1	664	2	E69657 This is a DE line.	1.73e-08
13	173	14.4	650	2	A64572 This is a DE line.	1.29e-07
14	160	13.4	677	1	SYECMT This is a DE line.	3.19e-06
15	158	13.2	810	2	S75931 This is a DE line.	5.21e-06
16	138	11.5	246	2	E69230 This is a DE line.	6.32e-04
17	128	10.7	764	2	B64570 This is a DE line.	6.47e-03
18	108	9.0	201	2	A69999 This is a DE line.	5.67e-01
19	106	8.8	110	2	E69607 This is a DE line.	8.73e-01
20	106	8.8	795	1	SYECFB This is a DE line.	8.73e-01
21	105	8.8	805	1	S73374 This is a DE line.	1.08e+00
22	102	8.5	137	2	S30117 This is a DE line.	2.05e+00
23	100	8.3	336	2	D69496 This is a DE line.	3.12e+00

ALIGNMENTS

RESULT ID	1	STANDARD	PRT	310 AA.
XX	A55053	xxxxxx		
XX	01-JAN-1900			
XX		This is a DE line.		
XX				
CC	A:Accession: A44032			
CC	A:Molecule type: protein			
CC	A:Residues: 145-158, 'X', 160-164 <KA2>			
CC	A:Experimental source: methylcholanthrene A fibrosarcoma cells			
CC	A>Note: sequence extracted from NCBI backbone (NCBIR:115676)			
CC	SEQUENCE 310 AA; 33997 MW; 487802 CN;			
DB	148 IDASRLDIRIGCTVAKKHPDASLYEEVDVGEAPRTVSGLVNHPLEQNRMYVL 207			
OY	3 VIPSRDIRVGKIIIVKHPDASLYEKGAEAPRTVSGLVQFVKELQRLVYV 62			
DB	208 LCNLRPAKRGVLSQAMWCAS--SPK-VETLAPRNGSVPRDRTTFDAFP-GEPRKELN 263			
OY	63 LCNLRPAKRGVLSQAMWCASIEGINROVEPLDPASAPAEHVFVKGEGQDEELK 122			
DB	264 PKRKIMEQIOPDLHTNACVATYKGAPEVK-GKVCRA 301			
OY	123 PKKVFELQADFKEISCEIAQMKOTNFMTKIGSISCKS 161			
RESULT ID	2	STANDARD	PRT	310 AA.
XX	B55053	xxxxxx		
XX	01-JAN-1900			
XX		This is a DE line.		
XX				
CC	A:Title: Characterization of a novel tumor-derived cytokine. Endothelial-monocy			
CC	A:Reference number: A55053			
CC	A:Accession: B55053			
CC	A>Status: preliminary			
CC	A:Molecule type: mRNA			

CC C:Genetics:
CC A:Gene: mets
CC C:Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis
SQ SEQUENCE 649 AA; 74354 MW; 2141119 CN;

Query Match	15.7%	Score 188;	DB 2;	Length 649;
Best Local Similarity	37.8%;	Pred. No. 2.95e-09;		
Matches	31;	Conservative	23;	Mismatches 27; Indels 1; Gaps 1.

Db 551 AKVDLRVAEYVQPERKMNADKLKQLDLDG-GEKQVVISGAEFFYPEELICKKVICVAA 600
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 6 SRLDIRVGKLTVEKHPDADSLYEKIDVGEAEPRTVSGVQFVPEKELQDRLVWVICN 65

```

Db      610 LKPAKLRGEWSEGMILAGSGG 631
      ||| |::| |::|:| ::|
Qy      66 LKPQKMRGVESQGMILCASTEG 87

```

RESULT	11	STANDARD;	PRT;	664	AA
ID	S66067				
XY					

XX
DT
XX

01-JAN-1900

DE This is a DE line.

CC A:Cross-references: EMBL:ID26183; NID:g467326; PID:d1005815; PID:g467422
CC A>Note: the nucleotide sequence was submitted to the EMBL Data Library,
CC Genebank:

CC C;keywords: ligase
SQ SEQUENCE 664 AA; 76187 MW; 2333297 CN;

Query Match	15.1%;	Score 181;	DB 2;	Length 664;
Best Local Similarity	39.5%;	Pred. No. 1.73e-08;		
Matches	34;	Conservative	18;	Mismatches 31;
				Indels 3;
				Gaps 3;

```
D6      568 VELRAEVEIEAEPPVKADRLTKQLDLG-FEKROYVSGINKHYTPELVLGKLYCVT-NL 62
        ::::: :| | | | | :||| : :| : :||| :|
OY      8 LDIRVKIILYEKHPRDLSLYEKKIDVGAEAPRTIVSSGLV-QFVPKEELQDRLLVVTLGNL 66
```

```

Db      626 KPVKLRGELSQGMILLAGEDVLRV 65
        |||:| | | | : :|:|
Qy      67 KPQKMRGVESQGMILLCASIEGINRV 92

```

RESULT	12	STANDARD;	PRT;	664	AA
ID	E69657				
XX					

XX
DT 01-JAN-1900
XX

DE This is a DE line.
XX

```
CC A:Experimental source: strain 168
CC C:Genetics:
CC A:Gene: mets
CC C:Keywords: am[noacv]-tRNA synthetase
```

Query Match	15.18;	Score 181;	DB 2;	length 664;
Best local Similarity	39.5%;	Pred NO 1	73e-08.	
SEQUENCE	664 AA;	7618/ MW;	233329/ CN;	

Db 568 VELRYAEVIEAEPPVKADRLIKLQLDLG-FEKQVWSGIAKHYPPEELVGKKLYCVT-NL 6245

626 KPVKLRGELSGMILAGEADSVLKVV 651
QY 8 LDIRVGKIIIVEKEHPDADSLIVEKIDVGEAEPRRTWSGLV-QFVPREELQDRLVYVLCNL 66

67 K P Q K M R G V E S Q G M L T C A S I E G I N R Q V 92

RESULT	13
ID A64572	STANDARD; PRT; 650 AA

DT 01-JAN-1900

DE This is a DE line.

CC A:Title: The complete genome sequence of the gastric pathogen *Helicobacter*
CC A:Authors: Hayes, W.S.; Borodovsky, M.; Karpik, P.D.; Smith, H.O.; Fraser, C.M.;
CC A:Reference number: 264570. With: 07394457

CC A: Accession: A04572
CC A: Status: preliminary

CC A;Residues: 1-650 <

CC C;Keywords: ligase

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Query Match	14.48;	Score 173;	DB 2;	Length 650;
Best Local Similarity	35.58;	Pred. No. 1.29e-07;		
Matches	27;	Conservative	23;	Mismatches 24; Indels 2; Gaps 2

D_B 552 KVEIVGLILEAQRIEKSNKLRLKVDIGEGRLLQIISGALDIYP-ESLYGMWCYVAN 610
:::||| : : |::|| : ::|| : : || : :
QY 7 RLDIRVGKIITVEKHBDADSLYEKIDVGEAEPRVTVSGL-VQFPKEELDRLVVYLGN 65

```

Db      611 LKPAKLMGEMSEGMIL 626
      ||| | : | | : || : |
QY      66 LKPKMRGVESQGMILL 81

```

RESULT	14	STANDARD;	PRT;	677	AA
ID	SYECMT				
yy					

XX	01-JAN-1900
DT	
XY	

DE This is a DE line.
XX
CC A;Contents: annotation; X-ray crystallography, 2.5 angstroms
C;Concise:

```
CC A;Map position: 46 mkr
CC C;Function:
CC A;Description: catalyz
```

CC C;Superfamily: methionine--trna ligase
CC C;Keywords: aminocacyl-CRNA synthetase; ATP; homodimer; liase; metalloprotein;
CC E-3-677;Product: methionine--trna ligase #native predicted kDa
CC

CC F:62-66, 137-157, 335
CC F:452-468 / Position: 3

```
CC F;16,22,53,336/Active site: Tyr, His, Asp, Lys #status predicted
CC F;146,149,159,162/Binding site: zinc (Cys) #status experimental
SQ SEQUENCE 677 AA; 76254 MW; 2257999 CN;
```

Query Match	13.4%;	Score 160;	DB 1;	Length 677;
Best Local Similarity	33.7%;	Pred. No. 3.19e-06;		
Matches	35;	Conservative	26;	Mismatches 37;
			Indels	6;
			Gaps	6

```
QY      6 SRDLIRVGKIIIVKEKHPDADSLVEKIDVGEAPRTVVSGL-VGVYPKELQDRILVVLIC 64
```

```
Db      636 NLAPRKMRFGI-SEGVMMAAG-PG-GKDIFLLSPDAGAKPGHQV 670
        |||::||| |::|||:: |:: |:: |||:: |||:: |
```

[illegible]

	RESULT	15		
ID	S75931	STANDARD;	PRT;	810 AA.
XX	xxxxxx			
AC				
XX				
DY	01-JAN-1900			
XX				
DE	This is a DE line.			
XX				
CC	A:Accession: S75931			
CC	A>Status: preliminary			
CC	A:Molecule type: DNA			
CC	A:Residues: 1-810 <KAN>			
CC	A:Cross-references: EMBL:D90913; NID:g1653348; PTD:d1019123; PID:g1653476			
CC	A>Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996			
SQ	SEQUENCE 810 AA; 87887 MW; 3201926 CN;			
	Query Match	13.2%;	Score 158;	DB 2; Length 810;
	Best Local Similarity	40.7%;	Pred. No. 5.2E+06;	
Matches	37; Conservative	20; Mismatches	25; Indels	9; Gaps 6;
Dy	45 VILGVIVSEKHPNADKLSVCVVDIGTEEPSTIYGCAAVRADIIIVPATLGSLPKYDL 104 :: :: : :: : : : :			
OY	10 IRVGRIITVEKPDDSDLVVERIKIDGAEPRTVSGL--VQ---FVPKEELDRLYVV-L 63 :: : : : :::::			
Dd	105 -KKIPAKIRGVSSGM-I-CSIAELGIHRSKESE 133 :: : : : :::::			
OY	64 CMLRPGMRKGVSOGMILCASTIE-GINROYE 93			

Search completed: Sat Apr 18 02:05:28 1998
Job time : 54 secs.

THIS PAGE BLANK (USPTO)